

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 03:28:25 ; Search time 1190.53 Seconds  
(without alignments)  
6041.430 Million cell updates/sec

Title: US-09-521-640-2  
Perfect score: 465  
Sequence: 1 ggggagatcgtgagttcac.....ccctttgttcaaacacn 465

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Seq--ned: 1344157 seqs, 7733874588 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_em.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pl1.\*
- 13: gb\_pl2.\*
- 14: gb\_pl3.\*
- 15: gb\_pl4.\*
- 16: em\_ba1.\*
- 17: em\_ba2.\*
- 18: em\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_or.\*
- 43: em\_or.\*

- 44: em\_ov.\*
- 45: em\_pat.\*
- 46: em\_ph.\*
- 47: em\_pl.\*
- 48: em\_ro.\*
- 49: em\_sts.\*
- 50: em\_sy.\*
- 51: em\_un.\*
- 52: em\_v1.\*
- 53: gb\_sts1.\*
- 54: gb\_sts2.\*
- 55: gb\_sts3.\*
- 56: gb\_sy.\*
- 57: gb\_un.\*
- 58: gb\_v1.\*
- 59: gb\_v12.\*
- 60: gb\_htg1.\*
- 61: gb\_htg2.\*
- 62: gb\_htg3.\*
- 63: gb\_htg4.\*
- 64: gb\_htg5.\*
- 65: gb\_htg6.\*
- 66: gb\_htg7.\*
- 67: gb\_htg8.\*
- 68: gb\_htg9.\*
- 69: gb\_htg10.\*
- 70: gb\_htg11.\*
- 71: gb\_htg12.\*
- 72: gb\_htg13.\*
- 73: gb\_htg14.\*
- 74: gb\_htg15.\*
- 75: gb\_htg16.\*
- 76: gb\_htg17.\*
- 77: gb\_htg18.\*
- 78: gb\_htg19.\*
- 79: gb\_htg20.\*
- 80: gb\_htg21.\*
- 81: gb\_htg22.\*
- 82: gb\_htg23.\*
- 83: gb\_htg24.\*
- 84: gb\_htg25.\*
- 85: gb\_pr1.\*
- 86: gb\_pr2.\*
- 87: gb\_pr3.\*
- 88: gb\_pr4.\*
- 89: gb\_pr5.\*
- 90: gb\_pr6.\*
- 91: gb\_pr7.\*
- 92: gb\_pr8.\*
- 93: gb\_pr9.\*
- 94: gb\_rol.\*
- 95: gb\_rol2.\*
- 96: gb\_in4.\*
- 97: gb\_pr10.\*
- 98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score Match	Length	ID	Description
No matches found				

Search completed: November 7, 2001, 04:50:11  
Job time: 4906 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 03:48:16 ; Search time 129.24 Seconds  
(without alignments)  
2259.163 Million cell updates/sec

Title: US-09-521-640-2  
Perfect score: 465  
Sequence: 1 ggggagatcgtgagttcac.....ccctttgttcaaacacn 465

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Se.....ned: 730101 seqs, 313950809 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description
-----			

No matches found

Search completed: November 7, 2001, 04:52:32  
Job time: 3856 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 03:27:20 ; Search time 1162.82 Seconds  
(without alignments)  
3780.101 Million cell updates/sec

Title: US-09-521-640-2  
Perfect score: 465  
Sequence: 1 ggggagtagtgagttcac.....ccctctgttcaaaacacn 465

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Seq. ned: 10228115 seqs, 4726426750 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est25:\*
- 26: gb\_est26:\*
- 27: gb\_est27:\*
- 28: gb\_est28:\*
- 29: gb\_est29:\*
- 30: gb\_est30:\*
- 31: gb\_est31:\*
- 32: gb\_est32:\*
- 33: gb\_est33:\*
- 34: gb\_est34:\*
- 35: gb\_est35:\*
- 36: gb\_est36:\*
- 37: gb\_est37:\*
- 38: gb\_est38:\*
- 39: gb\_est39:\*
- 40: gb\_est40:\*
- 41: gb\_est41:\*
- 42: gb\_est42:\*
- 43: gb\_est43:\*
- 44: gb\_est44:\*
- 45: gb\_est45:\*
- 46: gb\_est46:\*
- 47: gb\_est47:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est33:\*
- 111: gb\_est34:\*
- 112: gb\_est35:\*
- 113: gb\_est36:\*
- 114: gb\_est37:\*
- 115: gb\_est38:\*
- 116: gb\_est39:\*
- 117: gb\_est40:\*
- 118: gb\_est41:\*
- 119: gb\_est42:\*
- 120: gb\_est43:\*
- 121: gb\_est44:\*
- 122: gb\_est45:\*
- 123: gb\_est46:\*
- 124: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.							
-----	-----	-----	-----	-----	-----	-----	-----

No matches found

Search completed: November 7, 2001, 04:30:08  
Job time: 3768 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 03:24:30 ; Search time 1194.37 Seconds  
(without alignments)  
6022.006 Million cell updates/sec

Title: US-09-521-640-2  
Perfect score: 465  
Sequence: 1 ggggagatcgtaggttcaac.....cccttctgttcaaacacn 465

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Seq..hed: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_on.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_on.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_v1.\*  
53: yb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_v11.\*  
59: gb\_v12.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: yb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_ro2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Queried	Length	DB	ID	Description
1	54.8	11	1719	65	AC018552	AC018552 Homo sapi
2	52.6	11	7218	10	I66494	I66494 Sequence 14
C 3	37.2	8	143786	6	CELY71H2AM	AC024859 Caenorhab
C 4	37.2	8	48960	60	AC006896	AC006896 Caenorhab
C 5	36.8	7	485	8	AF309412	AF309412 Oncorhync
C 6	36.4	7	215283	62	AC011966	AC011966 Homo sapi
C 7	36.2	7	40552	4	AC084469	AC084469 Caenorhab
C 8	36	7	174296	63	AC013805	AC013805 Homo sapi

```

 9  36  7.7 334796 70 AC026673 Homo sapi
10 35.8 7.7 132449 90 AL365272 Human DNA
11 35.8 7.7 171350 72 AC058816 Homo sapi
12 35.8 7.7 172307 81 AC044842 Homo sapi
13 35.8 7.7 182529 71 AC048962 Homo sapi
14 35.2 7.6 180538 88 AC073897 Homo sapi
15 35 7.5 160069 83 AP001977 Homo sapi
16 35 7.5 160800 83 AP002011 Homo sapi
17 35 7.5 183674 64 AC016996 Homo sapi
18 35 7.5 189078 83 AP001982 Homo sapi
19 34.8 7.5 16121 6 CEF40F11 273426 Caenorhabdi
20 34.8 7.5 85953 83 CEH04L24 297189 Caenorhabdi
21 34.8 7.5 107244 83 CER24F12 AL022277 Caenorhab
22 34.8 7.5 131318 83 CEY42E10 Z99944 Caenorhabdi
23 34.8 7.5 169689 72 AC051616 Mus muscu
24 34.6 7.4 1246 89 AF244570 AF244570 Homo sapi
25 34.6 7.4 1480 89 AF244571 AF244571 Homo sapi
26 34.6 7.4 1745 13 AF168859 Mayaca au
27 34.6 7.4 186647 69 AC025573 Homo sapi
28 34.6 7.4 32488 6 CELT12B5 AF100307 caenorhab
29 34.4 7.4 79531 90 AL390037 AL390037 Human DNA
30 34.2 7.4 73006 62 AC011388 Homo sapi
31 34.2 7.4 104271 60 AC008472 AC008472 Homo sapi
32 34.2 7.4 129389 92 HS526114 Z82214 Human DNA S
33 34.2 7.4 209336 74 AC073116 AC073116 Homo sapi
34 34.2 7.4 298575 84 HSB22484 AC080244 Homo sapi
35 33.8 7.3 1933 6 FPA224788 AJ224788 Flabellin
36 33.8 7.3 1941 6 ESA224786 AJ224786 Eubranchu
37 33.8 7.3 1969 5 AF249198 AF249198 Flabellin
38 33.8 7.3 2087 6 MLA224784 AJ224784 Melibe le
39 33.6 7.2 164399 74 AC073321 AC073321 Homo sapi
40 33.6 7.2 7764 89 AF225896 AF225896 Homo sapi
41 33.6 7.2 17245 7 OCAJ1588 AJ001588 Oryctolag
42 33.6 7.2 62043 90 AL161893 AL161893 Human DNA
43 33.6 7.2 138039 64 AC016786 AC016786 Homo sapi
44 33.6 7.2 152696 69 AC025195 AC025195 Homo sapi
45 33.6 7.2 153279 65 AC018395 AC018395 Homo sapi
```

## ALIGNMENTS

```

RESULT 1
AC018552 AC018552 171919 bp DNA HTG 26-SEP-2000
LOCUS Homo sapiens chromosome 16 clone RP11-405F3, WORKING DRAFT
DEFINITION SEQUENCE, 19 unordered pieces.
ACCESSION AC018552
VERSION AC018552.4 GI:10305141
KEYWORDS HTG; HTGS_PRAISE1; HTGS_DRAFT.
Src "E human.
\NISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171919)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171919)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 26, 2000 this sequence version replaced gi:8576099.
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-11_405F3
-----
```

```

Summary Statistics
Consensus quality: 130458 bases at least Q40
Consensus quality: 142777 bases at least Q30
Consensus quality: 149702 bases at least Q20
Estimated insert size: 171000; agarose-fp estimation
Estimated insert size: 170119; sum-of-contigs estimation
Quality coverage: 4.73 in Q20 bases; agarose-fp estimation
Quality coverage: 4.76 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1194: contig of 1194 bp in length
* 1294: gap of unknown length
* 2441: contig of 1147 bp in length
* 2541: gap of unknown length
* 3628: contig of 1087 bp in length
* 3728: gap of unknown length
* 3729: contig of 1094 bp in length
* 4822: gap of unknown length
* 6068: contig of 1146 bp in length
* 6168: gap of unknown length
* 7451: contig of 1283 bp in length
* 7551: gap of unknown length
* 8563: contig of 1012 bp in length
* 8663: gap of unknown length
* 10826: contig of 2163 bp in length
* 10926: gap of unknown length
* 12392: contig of 1366 bp in length
* 12393: gap of unknown length
* 13481: contig of 1089 bp in length
* 13581: gap of unknown length
* 14780: contig of 1199 bp in length
* 14880: gap of unknown length
* 18391: contig of 3511 bp in length
* 18491: gap of unknown length
* 21416: contig of 2925 bp in length
* 21516: gap of unknown length
* 29200: contig of 7684 bp in length
* 29300: gap of unknown length
* 43025: contig of 13725 bp in length
* 43125: gap of unknown length
* 55543: contig of 12418 bp in length
* 55643: gap of unknown length
* 70587: contig of 14944 bp in length
* 70687: gap of unknown length
* 94927: contig of 24240 bp in length
* 94928: gap of unknown length
* 95028: 171919: contig of 76892 bp in length.
FEATURES
    source
        1..171919
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="RP11-405F3"
            /clone_lib="RPCI human BAC library 11"
BASE COUNT 40913 a 42871 c 43420 g 42804 t 1911 others
ORIGIN
Query Match 11.8% Score 54.8; DB 65; Length 171919;
Best Local Similarity 61.0%; Pred No. 1.3e-05;
Matches 89; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
Qy 113 gagaacacgcgtaccgcgtactctcttcacacacccctatcgccggtatctca 172
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10105 GAAACCCCTGCGTACCAACTAATTCCTGGGACTTCCTTTTACCACACTCGTAAGCAA 10164
Qy 173 atgaccaccatcgctctcccaaccatgccaaactaatggcaatggccctgatcggtttctc 232
```



Db	10165	AAGCCACGATCCCTTCCACACATCCACCTGAGGCAATGCGCCCAACGATTTCTC	10224
Qy	233	ttaccattgtcggtattaccgcgtat	258
Db	10225	TTTACATTTGGGGATTTCACCATAT	10250
RESULT 2			
LOCUS	166494	I66494 7218 bp DNA	PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.		
ACCESSION	I66494		
VERSION	I66494.1	GI:2724471	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REMARKS	Unclassified.		
AUTHORS	Donner, F., Scheffinger, F. and Falkner, F. Gunter.		
TITLE	Recombinant fowlpox virus		
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;		
FEATURES	Location/Qualifiers		
source	1..7218		
BASE COUNT	1944 a 1491 c 1486 g 1929 t	368 others	
ORIGIN			
Query Match	11.3%; Score 52.6; DB 10; Length 7218;		
Best Local Similarity	6.5%; Pred No. 4.8e-05;		
Matches	28; Conservative 220; Mismatches 180; Indels 0; Gaps 0;		
Qy	37	agactcgctgaagaccctgcgtccagcgaatcccttatgagcgttaactgaactgcgcg	96
Db	1036	AGTTGGCTGCAGGTCGAGGAGCTTGCATTTTTTTTTTTTTTTTTTTTTTTTTT	1095
Qy	97	cgttacacgaggactgagaacactgcgtacccgctaactctcttcacacatccctatcg	156
Db	1096	YTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1155
Qy	157	cactggcggtatcctaagaccaccatcgcttccacacatgccaactaaggcaatgga	216
Db	1156	YTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1215
Q	217	cctgacgattctcttaccattgctggtatcccgctatgacgcttccatcaacagag	276
Db	1216	YTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1275
Qy	277	gatgccttagtaccctaccacaccocgtgacacaccccttgaccttaataaaacct	336
Db	1276	YTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1335
Qy	337	atcttctcattcggaaataactctctctcttcagcaaatgaagaccctgctctg	396
Db	1336	YTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1395
Qy	397	tccttggcggccttttgcacgccttttggactgtttatcaaaaaccctcttctg	456
Db	1396	YTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1455
Qy	457	caaaacac	464
Db	1456	TTAACTAC	1463
RESULT 3			
LOCUS	CEL71H2AM/c		
DEFINITION	Caenorhabditis elegans cosmid Y71H2AM, complete sequence.		
ACCESSION	AC024859		
VERSION	AC024859.1	GI:7140422	
KEYWORDS	HTG.		

SOURCE	ORGANISM
REFERENCE	Caenorhabditis elegans.
AUTHORS	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
	1 (bases 1 to 143786)
	The C. elegans Genome Sequencing Consortium, Washington University Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre, Hinxton, U.K., C.
TITLE	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)
JOURNAL	99069613
MEDLINE	2 (bases 1 to 143786)
REFERENCE	Waterston, R.H.
AUTHORS	Direct Submission
TITLE	Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
JOURNAL	3 (bases 1 to 143786)
	Waterston, R.
REFERENCE	Direct Submission
AUTHORS	Submitted (06-APR-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE	Submitted by:
JOURNAL	Genome Sequencing Center
COMMENT	Department of Genetics, Washington University, St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
NOTICE:	This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.
NOTES:	
Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).	
FEATURES	source
	1..143786
	/organism="Caenorhabditis elegans"
	/strain="Bristol N2"
	/db_xref="taxon:6239"
	/chromosome="III"
	/clone="CEL71H2AM"
	complement(2225..8816)
gene	/gene="Y71H2AM.9"
	complement(join(2225..2326,2429..2493,3402..3513,3572..3696,5645..5789,6811..7111,7191..7389,7443..7489,8689..8816))
CDS	/gene="Y71H2AM.9"
	/codon_start=1
	/evidence=not_experimental
	/product="Hypothetical protein Y71H2AM.9"
	/protein_id="AAK29970.1"
	/db_xref="GI:13559766"
	/translation="MKSMLKMNPFEGFKGFASFEFLAKKLVKNSKKIKIKIITWAGKIPMSLVISNVNTPFRKNPILMVSVRCLSSNQPTDDHHSKRKLKLFKVAQLS KRSYENDKIDKLUSAEEFGLKESDLELPSSKPSVLDSHGHGKSNHQLTFLTEVYH ALKHGSTEALTAELMRPLVAMRKELTDAERMNRVETTNQHKGDADVVAFTLN DMAMFTAYLTGSKSLFAEATHSAMDCNQILLIGIRYSKPNDFLLFPYGYGNM YVTSLSGGGIMAFGGGLSMYHGISGLLHPEPLEPTYAYVALFMSLCFOGTSAITA YVFAKAKKANISIMNVVTSADPSLVNVLLEDTAAVTGVIAGLCAISLSFLNSPI YVAGSIVIGALLGS"
	1..14531
gene	/gene="Y71H2AM.8"
	join(9613..9658,9753..9909,10763..10879,12629..13172,14056..14418,14490..14531)
CDS	/gene="Y71H2AM.8"

gene  
CDS  
/codon\_start=1  
/evidence=not\_experimental  
/product="Hypothetical protein Y71H2AM.8"  
/protein\_id="AAK29971.1"  
/db\_xref="GI:13559767"  
/translation="MARSPTAELEDDQILGKIATLFEETLSKQPPSNIVELWKEHVEIR  
RLNDVSIQAKLGDTDERLANSRRDASIKTFILAWDGVGIARNVITIGSTKTAGLS  
LQATGPPIKAVPRPHAMITDLAKKSLKAKFEKDPVIGGMVNGVGLALACQ  
WIAKESKWSYISLPTTPTFLFYEEQLQKPSIFEEALFYRTISRCQCYFL  
LQANKIYEAAQRKDAARNMETPIFYNPVFNANFTPKLYFWAGVVTTRVNVPS  
ENQYEGDNPVIIIPALIPVDMANHENVLTDLTEPLEDLVCSPEEACAVITSHCDV  
KAGNEVIFYCSCRSGEHLHNGFVPIYHGREDVLKIGIPKTKDXTLDAKKLIQKE  
KVVYICAGNIFVDVILNCKNMRISLKIRFC"  
/complement(16532..23598)  
/gene="Y71H2AM.10"  
complement(join(16532..16726,17546..17679,18063..18187,  
18239..18398,19350..19840,20512..21131,21447..21702,  
21749..21883,22158..22268,22316..22553,23146..23251,  
23491..23598))  
/gene="Y71H2AM.10"  
/note="coded for by C. elegans cDNA yk285g2.5; coded for  
by C. elegans cDNA yk169a12.5; coded for by C. elegans  
cDNA yk285g2.3; coded for by C. elegans cDNA yk169a12.3"  
/codon\_start=1  
/product="Hypothetical protein Y71H2AM.10"  
/protein\_id="AAK29969.1"  
/db\_xref="GI:13559765"  
/translation="MATVYSADLSRRRVFEEIPLHCAVRVRIITEKVVTRPVVDLL  
AVPDESIRHVGATCVHAQLTSSTATELGKCYLNHRHEDOKSCVIRIPVPSFPPVD  
ONRYSVLSVSYVSEKCDONEDHLDLPOHLIEVNSRIAKQKIDWLANATETSVLLSTAS  
QAFSONQMTLFLHVSWANETQPEIRLWDSMSIETVPTSSNWTIRVSSASRPD  
FLYSLTQPKRMTGFNDINVAIILIKMVSDTAIKDDVILHWHVILGPKSEPPDDH  
KATFVSIADEAAVTVKPCGRQYITFHAKKTSRHRKELANLAVISGIVYH  
SRLAFITISAKAEDVTQSHCISSDANIILKPTSCSVYLDGSESSGSDAAYHAT  
VRYTTHVSFVPPKPLKIKWSSSTLSTIKNKGFWRLPLGGGVKRSRAARQPA  
CVNRQFVSLATLIEDIKTDGLYLSHKSITLEDVNIHVTNLTQISNRTVANK  
FYERAKVIGNVGLAKLIVRNAAKMDLVSENISVQNKVESTTGLSARPCDTSFRI  
LPILPFAFFVEIAHKSLLKLYQCCIFASVSDATWEPLNDLSDSYPFMSAHS  
NERALVSHASKVHVTAIDENWPLPVVEISLQSAQASTINGASPAALAVTVLNP  
IKINSPSATDLSSESTTAPSDSPSPFPHFVLTIIGLIILFLIFSVRRSAF  
KGYQLVFPFSSRLSSSGNSRQEBETNEWMLSQOPPSSTISGSGNKSSTAEROS  
SNGDDPSRTSISYHGEISVFIAFSGQNVVYVQSSRHPRYTLVDSNHDNLARIVPKE  
DRWTTGGEQFHTWQKRGSGGGRMEAPRESIA"  
/complement(27014..33736)  
/gene="Y71H2AM.11"  
complement(join(27014..27225,27780..27833,27889..28391,  
29951..30243,32146..32213,33337..33442,33535..33658,  
33711..33736))  
/gene="Y71H2AM.11"  
/note="coded for by C. elegans cDNA yk646h6.5; coded for  
by C. elegans cDNA yk646h6.3"  
/codon\_start=1  
/product="Hypothetical protein Y71H2AM.11"  
/protein\_id="AAK29983.1"  
/db\_xref="GI:13559779"  
/translation="MDQIDAVFKSIDERQDEFIDLLRESVAIQSVADPARRGDCVRM  
SEWARDQKTLGVTSWELGQQLPSGEQLPLPPAVGVYGRDKSRKTLIYGLHDV  
QPAEKEDQNTPELTIDIGLFGRGSTDGKGVIAWIAVLQVLTIGLDLPINIRK  
VLEWDESSGLDKGBDNKIDSIVTFCISIDNYWLGKRNKPCLTPEALNDLWVMS  
QLVTVDGQIILPGIAELVAPLTKDEDEIYEKIDFCVDPRFKNETGSHGLMSDNKNLLM  
NWRYPSSLHGVCAGSQPQAKTVIPAKVVGKFSRLRVPHMTPEATDKLVNSYLDLS  
WAKRSPNFTKVTSGHGMPPWADFDRANFSAGSRAIERLENFRRKHSKRLNPPPKL  
YGMTPDFREGGISPVTLTIQDLTKSPVMLPLPIGASDDMAHSQNEKNIRDNFVKGKV  
LAAYLIFELAAA"  
35242..40205  
/gene="Y71H2AM.7"  
join(35242..35313,35416..35535,37225..37334,38048..38494,  
39905..40097,40170..40205)  
/gene="Y71H2AM.7"  
/codon\_start=1  
/evidence=not\_experimental  
/product="Hypothetical protein Y71H2AM.7"  
/protein\_id="AAK29972.1"

gene

CDS

gene  
CDS  
/db\_xref="GI:13559768"  
/translation="MAVDIRGVYIGNSMFILOKFVNFKLIPNFSERQNAANQSDPFDM  
KPKLYEVVFTVCRILRPNLSRRQVNTNRKIGLPTYSIVASESQRFVKLIGFKM  
PDSLDCACEMALKVLTFTTKRGMIDBEKYNLQWHTLIVLQVCFINHIIEYERFIRK  
CPALCRTEERHGFSLQFPHGCRGVVSRPMQROVLTITVYLFASFCLSFRIQDF  
RFPLIKINKYFLQMLKQFPSPKTKMYSVGKORLNIISFKWDIQLAANTVOTAYILL  
GTSOIANVSVIINLRCDNAIVPELPKQSIIEELACAKNESIPECSTSS"  
/complement(40206..42546)  
/gene="Y71H2AM.12"  
complement(join(40206..40261,41089..41240,41746..41854,  
41904..42012,42332..42444,42498..42546))  
/gene="Y71H2AM.12"  
/note="coded for by C. elegans cDNA yk301b4.5; coded for  
by C. elegans cDNA yk234f4.5; coded for by C. elegans cDNA  
yk301b4.3"  
/codon\_start=1  
/product="Hypothetical protein Y71H2AM.12"  
/protein\_id="AAK29982.1"  
/db\_xref="GI:13559778"  
/translation="MDTIKTVKVVVVGESGAGKTALLTCFLDNTFETDPLTIGIDFK  
HKIVQLNDGOSIRLQMDTAGERFQRLAPAYIRSAVALLVIDLSDENCYVHLIRNK  
GIIDKNKSDFTTIIVGNKHDLVSEKSRPLTAIIRETNDEYIETSAKMRNIIKLF5  
SVACRPPHEHTSOIILLNEPRPVESATKRCCQRW"  
complement(42822..48202)  
/gene="Y71H2AM.13"  
complement(join(42822..43008,43641..43770,44200..44526,  
44910..45066,45113..45297,46544..46943,47007..47099,  
47964..48023,48089..48202))  
/gene="Y71H2AM.13"  
/note="coded for by C. elegans cDNA yk75h9.5; coded for by  
C. elegans cDNA yk647e3.5; coded for by C. elegans cDNA  
yk233h4.5; coded for by C. elegans cDNA yk513a6.5; coded  
for by C. elegans cDNA yk100g10.5; coded for by C. elegans  
cDNA yk65c4.5; coded for by C. elegans cDNA yk169d11.5;  
coded for by C. elegans cDNA yk513a6.3; coded for by C.  
elegans cDNA yk169d11.3; coded for by C. elegans cDNA  
yk233h4.3; coded for by C. elegans cDNA yk647e3.3; coded  
for by C. elegans cDNA yk75h9.3; coded for by C. elegans  
cDNA yk378b12.3"  
/codon\_start=1  
/product="Hypothetical protein Y71H2AM.13"  
/protein\_id="AAK29968.1"  
/db\_xref="GI:13559764"  
/translation="MGGVSKHTTTPHPVPPSRQVLTSGPIEGRLIHGEKQVDAFQ  
IPYAAPPLIGNRFPALPOPEKWEVRETSKFGAGIQKHVLSPKTSQSDNLTNI  
FTPYTPKNETGFVILYIHGGFVSDSAHRYGDMSCQHLVTKDVVVVTIQLRGL  
GFWTGTGSSIPDNVALHDMVPAKWKVNICLNGDPNNITLMQSGAGSVDFLSIS  
PVSRDLQKVIIPMGNGASCWAIIHPRPLNACRRAOEIGVDFGNTLDWVKLRLELPA  
DKFASALNMEAVDTKDPPELLIGPKYDNLFIPIPRVMELEKAPKPRLLGCAKSEGLV

Query Match 8.0%; Score 37.2; DB 6; Length 143786;  
Best Local Similarity 53.1%; Pred. NO. 2.8;  
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 317 ttggccttaataaaacctctttgtcatttcggaataaaaccttctgtctttcagca 376  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 79188 TTTTCCAAAAAACCACCAATTTTTCGAAATTTGAAAAATATTTTTCATTTTAAACA 79129  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 377 aatgaagaacctgctcngtctttgsgcggtctttgacgcgtctttgacctgtta 436  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 79128 AATTTCAAAAAAATCAATTTTTCGAAATTTTTCGAAACTATTTTCGATTTTTT 79069  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 437 tcaaaaaaacctctctgttcaaaaaa 463  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 79068 GAAAAAATCAATATTCTGAAAAA 79042  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
RESULT 4  
AC006896/c  
LOCUS  
DEFINITION Caenorhabditis elegans clone Y71H2X, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 15 unordered pieces.  
ACCESSION AC006896

gene

CDS

```

VERSION      AC006896.2  GI:4309903
KEYWORDS     HTG: HTGS_PHASE1
SOURCE       Caenorhabditis elegans.
ORGANISM     Caenorhabditis elegans.
REFERENCE    1 (bases 1 to 298960)
AUTHORS      Waterston,R.H.
TITLE        The sequence of Caenorhabditis elegans clone
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 298960)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (24-FEB-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
CO "NT
* On Mar 1, 1999 this sequence version replaced gi:4263452.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 3199: contig of 3199 bp in length
* 3200
* 3213: gap of unknown length
* 3214
* 5295: contig of 2082 bp in length
* 5296
* 5309: gap of unknown length
* 5310
* 8249: contig of 2940 bp in length
* 8250
* 8263: gap of unknown length
* 8264
* 11524: contig of 3261 bp in length
* 11525
* 11538: gap of unknown length
* 11539
* 15051: contig of 3513 bp in length
* 15052
* 15065: gap of unknown length
* 15066
* 21691: contig of 6626 bp in length
* 21692
* 21705: gap of unknown length
* 21706
* 29305: contig of 7600 bp in length
* 29306
* 29319: gap of unknown length
* 29320
* 39544: contig of 10222 bp in length
* 39545
* 39555: gap of unknown length
* 39556
* 57970: contig of 18415 bp in length
* 57971
* 57984: gap of unknown length
* 57985
* 68995: contig of 11911 bp in length
* 68996
* 68909: gap of unknown length
* 69110
* 90124: contig of 20201 bp in length
* 90111
* 113337: contig of 23213 bp in length
* 113338
* 113351: gap of unknown length
* 113352
* 154606: contig of 41255 bp in length
* 154607
* 154621: gap of unknown length
* 154622
* 264746: contig of 110126 bp in length
* 264747
* 264761: gap of unknown length
* 264762
* 298960: contig of 34200 bp in length.
*
* Location/Qualifiers
*     1..298960
*       /organism="Caenorhabditis elegans"
*       /db_xref="taxon:6239"
*       /clone="Y71H2x"
*
* BASE COUNT  96038 a 53465 c 53550 g 95711 t 196 others
* ORIGIN
*
* Query Match      8.0%; Score 37.2; DB 60; Length 298960;
* Best Local Similarity 53.1%; Pred. No. 2.9;
* Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
*
* QY 317 ttggccttaataaaacctatttgcatttcggaataaacatttcctgtcttcagca 376
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* Db 153844 TTTTCCCAAAAAAACCACCAATTTGAAATTTGGAATAATTTCCATTTTAAACA 153785
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* QY 377 aatgaagaacccctgctcgtccttttggcgcccttttgcacgcttttggacttgta 436
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 153784 AATTTCAAAAAAATCATTTTTTTTGTGAAATTTTTCGAAACTATTTTTCGATTTTTT 153725
QY 437 tcaaaaaaacctctctgttcaaaaca 463
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153724 GAAAAAAATCAATTTATTTCTGAAAAAA 153698
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AF309412/c AF309412 485 bp DNA VRT 24-OCT-2000
LOCUS      Oncorhynchus mykiss 18S ribosomal RNA gene, partial sequence.
DEFINITION
ACCESSION  AF309412
VERSION     AF309412.1  GI:10954038
KEYWORDS   rainbow trout.
SOURCE     Oncorhynchus mykiss
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei;
           Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE  1 (bases 1 to 485)
AUTHORS    Jones,I., Kille,P., Wigham,T. and Sweeney,G.E.
TITLE      Cloning and characterization of rainbow trout 18S rRNA
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 485)
AUTHORS    Jones,I., Kille,P., Wigham,T. and Sweeney,G.E.
TITLE      Direct Submission
JOURNAL    Submitted (28-SEP-2000) Department of Biosciences, Cardiff
           University, Museum Avenue, Cardiff CF1 3US, UK
FEATURES   source
           1..485
           /organism="Oncorhynchus mykiss"
           /db_xref="taxon:8022"
           <1..>485
           /product="18S ribosomal RNA"
BASE COUNT  119 a 117 c 144 g 105 t
ORIGIN
*
* Query Match      7.9%; Score 36.8; DB 8; Length 485;
* Best Local Similarity 53.5%; Pred. No. 2.6;
* Matches 99; Conservative 0; Mismatches 83; Indels 3; Gaps 1;
*
* QY 229 tctcttaccattgcgatttaccgcgtatggcgttcatcaacagagatgcctagta 288
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* Db 364 TCTCGTTTCGTTATCGGAATTAACGACAGACAAATCGCTCCACCACTAAGACGGCCATGCA 305
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* QY 289 tcacctaccaccccgctgaccaaaccttggcctttaaataaaacctatcttgcatt 348
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* Db 304 CCACCCACCACAGATC---GAGAAAGAGCTATCAATGTGTCATCTTCCGTGTCGG 248
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* QY 349 tcggaataaaccttctctgtcttccagcaaatgaagacccctcgtccttggcgcg 408
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* Db 247 GCCGGTGAGGTTTCCCGTGTGGTCAAAATTAAGCCGCACTGCCCACTCCTCGGTG 188
* ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
* QY 409 gccctt 413
* |||||
* Db 187 CCCTT 183

RESULT 6
AC011966/c AC011966 215283 bp DNA HTG 23-DEC-2000
LOCUS      Homo sapiens chromosome 15 clone RP11-285114 map 15, WORKING DRAFT
DEFINITION
ACCESSION  AC011966
VERSION     AC011966  GI:11990705
KEYWORDS   HTG; HLA; PIASE1; HTGS_DRAFT.
SOURCE     Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 215283)
AUTHORS    Birrell,M., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE	Homo sapiens chromosome 15, clone RP11-285F14
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 215283)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B., Brown, A., Castie, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Headfar, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lechoczky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE	JOURNAL	COMMENT
Direct Submission	Submitted (17-OCR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	On Dec 23, 2000 this sequence version replaced gi:7960359. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2605
Center clone name: 285-1-14
----- Summary Statistics

```

Sequencing vector: M13; M77815; 47% of reads  
 Sequencing vector: Plasmid; n/a; 53% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 193454 bases at least Q40  
 Consensus quality: 200496 bases at least Q30  
 Consensus quality: 205965 bases at least Q20  
 Insert size: 185000; agarose-fp  
 Insert size: 213083; sum-of-contigs  
 quality coverage: 11.3 in Q20 bases; agarose-fp

Quality coverage: 9.8 in Q20.

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	I	1012:	contig of 1012 bp	in length
*	1013	1112:	gap of 100 bp	
*	1113	2148:	contig of 1036 bp	in length
*	2149	2248:	gap of 100 bp	
*	2249	3454:	contig of 1206 bp	in length
*	3455	3554:	gap of 100 bp	
*	3555	4747:	contig of 1193 bp	in length
*	4748	4847:	gap of 100 bp	
*	4848	5856:	contig of 1009 bp	in length
*	5857	5956:	gap of 100 bp	
*	5957	7144:	contig of 1188 bp	in length
*	7145	7244:	gap of 100 bp	
*	7245	8271:	contig of 1027 bp	in length
*	8272	8371:	gap of 100 bp	
*	8372	9382:	contig of 1011 bp	in length
*	9383	9482:	gap of 100 bp	
*	9483	10621:	contig of 1139 bp	in length
*	10622	10721:	gap of 100 bp	
*	10722	11851:	contig of 1130 bp	in length
*	11852	11951:	gap of 100 bp	
*	11952	13017:	contig of 1066 bp	in length

7

*	13018	13117:	gap of	100 bp
*	13118	14592:	contig of	1475 bp in length
*	14593	14692:	gap of	100 bp
*	14693	15932:	contig of	1240 bp in length
*	15933	16032:	gap of	100 bp
*	16033	17854:	contig of	1822 bp in length
*	17855	17954:	gap of	100 bp
*	17955	19826:	contig of	1872 bp in length
*	19827	19926:	gap of	100 bp
*	19927	21064:	contig of	1138 bp in length
*	21065	21164:	gap of	100 bp
*	21165	23108:	contig of	1944 bp in length
*	23109	23208:	gap of	100 bp
*	23209	24764:	contig of	1556 bp in length
*	24765	24864:	gap of	100 bp
*	24865	26747:	contig of	1883 bp in length
*	26748	26847:	gap of	100 bp
*	26848	28606:	contig of	1759 bp in length
*	28607	28706:	gap of	100 bp
*	28707	30429:	contig of	1723 bp in length
*	30430	30529:	gap of	100 bp
*	30530	32780:	contig of	2251 bp in length
*	32781	32880:	gap of	100 bp
*	32881	215283:	contig of	192403 bp in length

```
FEATURES
  source
    Location/Qualifiers
      1..215283
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="15"
        /map="15"
        /clone="RP11-385I14"
        /clone.lib="RPCI-11 Human Male BAC"
      1..1012
        misc feature
```

`misc_feature`

```

1. .1012
|/note="assembly_fragment"
1113. .2148
|/note="assembly_fragment"
2249. .3454
|/note="assembly_fragment"
3555. .4747
|/note="assembly_fragment"
4848. .5956
|/note="assembly_fragment"
5957. .7144
|/note="assembly_fragment"
7245. .8271
|/note="assembly_fragment"
8372. .9382
|/note="assembly_fragment"
9483. .10621
|/note="assembly_fragment"
10722. .11851
|/note="assembly_fragment"
11952. .13017
|/note="assembly_fragment"
13118. .14592
|/note="assembly_fragment"
14893. .15932
|/note="assembly_fragment"
16033. .17854
|/note="assembly_fragment"
17955. .19826
|/note="assembly_fragment"
19927. .21064
|/note="assembly_fragment"
21165. .23108
|/note="assembly_fragment"
23209. .24764
|/note="assembly_fragment"
24865. .26747
|/note="assembly_fragment"
26848. .28606
|/note="assembly_fragment"
28707. .30429

```

```

misc_feature      /note="assembly_fragment"
30330..32780
/note="assembly_fragment"
32881..215283
/note="assembly_fragment"
60660 a 43707 c 46632 g 62078 t 2206 others
BASE COUNT
ORIGIN

Query Match      7.8%; Score 36.4; DB 62; Length 215283;
Best Local Similarity 61.1%; Pred. No. 5;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 331 aaacctatcttgcatttcgaataaactttctctgtcttccagcaaatgaagaacctc 390
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62995 ACACCTATCAACCTCTCTTATTATAAATCATCTCTTTTAGTGACTCAAAATCCACT 62836
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

C 391 gctcngtccttggcgggccctttgcaagctttt 425
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62835 GCCACTTCCTGGGTGAGAGCTCTCACACAACTTT 62801
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
AC084469
LOCUS      AC084469 40552 bp DNA INV 04-NOV-2000
DEFINITION Caenorhabditis briggsae cosmid CB045D24, complete sequence.
ACCESSION AC084469
VERSION AC084469.1 GI:11094919
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 40552)
TITLE Washington University Genome Sequencing Center.
JOURNAL The C. briggsae Genome Sequencing Project
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 40552)
Waterston,R.
Direct Submission
Submitted (04-NOV-2000) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspie@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES
source
1..40552
Location/Qualifiers
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="CB045D24"
BASE COUNT 12516 a 7553 c 7254 g 13229 t
ORIGIN

Query Match      7.8%; Score 36.2; DB 4; Length 40552;
Best Local Similarity 53.6%; Pred. No. 5.2;
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 323 ttaataaaacacctatcttgcatttcgaataaactttctctgtcttccagcaaatgaa 382
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3472 TTTAATGAAACCTTACTGTATGATGTTACGGTATTAATAATACCGTATTTCCCAATAGAA 3531
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 383 gaacctgtctcgtcttggcgggccctttgcaagctttttggacttttatcaaaa 442
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 3532 GCCCACAGAACCGTCATCTTTGAGCTCAACAATCTCGATTCTGCTGAAGATATCAAAA 3591

Qy 443 aaacctctctgttcaaaa 460
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3592 AGAGCCCAACTGATAAAA 3609
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AC013805/c
LOCUS      AC013805 174296 bp DNA HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-2003, WORKING DRAFT SEQUENCE, 36 unordered
pieces.
ACCESSION AC013805
VERSION AC013805.4 GI:7249021
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174296)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-2003
Unpublished
2 (bases 1 to 174296)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6715925.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3956
Center clone name: 20_O_3
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142071 bases at least Q40
Consensus quality: 155814 bases at least Q30
Consensus quality: 163016 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 170796; sum-of-contigs
Qual. coverage: 2.9 in Q20 bases; agarose-fp
Qual. coverage: 3.0 in Q20 bases; sum-of-contigs
-----
* This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of Ns. But the exact sizes of the gaps are unknown.
* This will be updated with the finished sequence
* as it is available and the accession number will
* be printed.

```



```

Db 13779 ATGTCACATATGTTTCACATATCTCTGGCCCTTGTGCAAAATGCTGTTCCCTCCATTTGGA 13720
Oy 407 gggcctttgcacgtcttttggactgttatcaaaaaaaccttc 451
Db 13719 CAGCTTTTCTTTCAGCTTGGCACTTGGAAATCCCTAAATCCATTC 13675

RESULT 9
AC026673 334796 bp DNA HTG 15-NOV-2000
LOCUS Homo sapiens chromosome 3 clone RP11-56B20, WORKING DRAFT SEQUENCE,
DEFINITION 70 unordered pieces.
ACCESSION AC026673
VERSION AC026673.12 GI:10047500
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SC "E" human.

UNIM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 334796)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amarutunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinch,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escott,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foister,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,K.R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu.L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 334796)
Worley,K.C.
Direct Submission
AUTHORS Submitted (23-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL On Sep 10, 2000 this sequence version replaced gi:9690210.
COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/

```

```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAOX
Center clone name: RP11-56B20
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 57% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 282951 bases at least Q40
Consensus quality: 301471 bases at least Q30
Consensus quality: 312551 bases at least Q20
Estimated insert size: 306588; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.1x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
consists of 70 contrigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contrigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1
50044: contrig of 50044 bp in length
50045 50144: gap of unknown length
50145 80526: contrig of 30382 bp in length
80527 80626: gap of unknown length
80627 106743: contrig of 26117 bp in length
106744 106843: gap of unknown length
106844 131778: contrig of 24935 bp in length
131779 131878: gap of unknown length
131879 145965: contrig of 14087 bp in length
145966 146065: gap of unknown length
146066 157223: contrig of 11458 bp in length
157224 157623: gap of unknown length
157624 168346: contrig of 10723 bp in length
168347 168446: gap of unknown length
168447 173627: contrig of 5181 bp in length
173628 173727: gap of unknown length
173728 179460: contrig of 5733 bp in length
179461 179560: gap of unknown length
179561 187040: contrig of 7480 bp in length
187041 187140: gap of unknown length
187141 193067: contrig of 5927 bp in length
193068 193167: gap of unknown length
193168 198818: contrig of 5651 bp in length
198819 198918: gap of unknown length
198919 204523: contrig of 5605 bp in length
204524 204623: gap of unknown length
204624 209428: contrig of 4805 bp in length
209429 209528: gap of unknown length
209529 212393: contrig of 2864 bp in length
212393 212493: gap of unknown length
212493 217105: contrig of 4613 bp in length
217106 217205: gap of unknown length
217206 222756: contrig of 5551 bp in length
222757 222856: gap of unknown length
222857 226203: contrig of 3347 bp in length
226203 230735: gap of unknown length
230735 230835: contrig of 4432 bp in length
230835 234307: gap of unknown length
234307 234407: contrig of 3472 bp in length
234407 238620: gap of unknown length
238620 238720: contrig of 4213 bp in length
238720 241740: gap of unknown length
241740 241840: contrig of 3020 bp in length
241840 244363: gap of unknown length
244363 244663: contrig of 2523 bp in length
244663 244663: gap of unknown length

```

```
* 244464 247852: contig of 3389 bp in length
* 247853 247952: gap of unknown length
* 247953 251204: contig of 3252 bp in length
* 251204 251304: gap of unknown length
* 251305 253244: contig of 2620 bp in length
* 253245 254024: gap of unknown length
* 254025 258689: contig of 4665 bp in length
* 258690 258789: gap of unknown length
* 258790 261355: contig of 2566 bp in length
* 261356 261455: gap of unknown length
* 261456 263987: contig of 2532 bp in length
* 263988 264087: gap of unknown length
* 264088 267491: contig of 3404 bp in length
* 267492 267591: gap of unknown length
* 267592 269625: contig of 2034 bp in length
* 269626 269725: gap of unknown length
* 269726 271829: contig of 2104 bp in length
* 271830 271929: gap of unknown length
* 271930 274961: contig of 3032 bp in length
* 274962 275061: gap of unknown length
* 275062 277332: contig of 2171 bp in length
* 277333 277332: gap of unknown length
* 277333 278978: contig of 1646 bp in length
* 278979 279078: gap of unknown length
* 279079 281944: contig of 2866 bp in length
* 281945 282044: gap of unknown length
* 282045 284703: contig of 2659 bp in length
* 284704 284803: gap of unknown length
* 284804 286499: contig of 1696 bp in length
* 286500 286599: gap of unknown length
* 286600 289303: contig of 2704 bp in length
* 289304 289403: gap of unknown length
* 289404 291372: contig of 1969 bp in length
* 291373 291472: gap of unknown length
* 291473 293293: contig of 1821 bp in length
* 293294 293393: gap of unknown length
* 293394 295064: contig of 1671 bp in length
* 295065 295164: gap of unknown length
* 295165 296870: contig of 1706 bp in length
* 296871 296970: gap of unknown length
* 296971 298568: contig of 1598 bp in length
* 298569 298668: gap of unknown length
* 298669 299734: contig of 1066 bp in length
* 299735 299834: gap of unknown length
* 299835 301632: contig of 1798 bp in length
* 301633 301732: gap of unknown length
* 301733 303247: contig of 1515 bp in length
* 303248 303347: gap of unknown length
* 303348 304560: contig of 1313 bp in length
* 304561 304760: gap of unknown length
* 304761 306454: contig of 1694 bp in length
* 306455 306554: gap of unknown length
* 306555 307630: contig of 1076 bp in length
* 307631 307730: gap of unknown length
* 307731 309334: contig of 1604 bp in length
* 309335 309434: gap of unknown length
* 309435 310603: contig of 1172 bp in length
* 310604 310706: gap of unknown length
* 310707 311844: contig of 1138 bp in length
* 311845 311944: gap of unknown length
* 311945 313346: contig of 1402 bp in length
* 313347 313446: gap of unknown length
* 313447 315156: contig of 1710 bp in length
* 315157 315256: gap of unknown length
```

Query Match 7.7%; Score 36; DB 70; Length 334796;  
Best Local Similarity 50.9%; Pred. No. 6.9;  
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy 287 tatcacccctaccacaccccgtagcaaaccttggccttttaataaaacacctatctgtgca 346
||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109665 TATCTCATTCCTCCACCCGTTTCACAAACTTAGTTCTGTGTAAGGGGAATCAGGCCTCA 109724
```

```
Qy 347 ttctcggaataaaactttctctgtcttttcagcaaatgaagaacccctgctgtcttgggc 406
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109725 ATGTCACATTATGTTACATATCTCTGGCCCTTTGCAATGCTGCTCCATTTGGA 109784
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 407 gggcctttggcagcgtttttggactgtttatcaaaaaaaccttc 451
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109785 CAGCTTTCTTTCAGCTTGGCACTTGGAAATCCCTAAATCCATTC 109829
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
LOCUS AL365272 132449 bp DNA PRI 23-DEC-2000
DEFINITION Human DNA sequence from clone RP11-328C17 on chromosome 6, complete
sequence.
ACCESSION AL365272
VERSION AL365272.20 GI:11991402
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Heath, P.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Dec 24, 2000 this sequence version replaced gi:11875910.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-328C17 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-328C17 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-328C17 is at 132449 in this
sequence. The true left end of clone RP11-233K4 is at 64893 in this
sequence. The true right end of clone RP3-416J7 is at 100 in this
sequence. The true right end of clone RP1-125A24 is at 61243 in
this sequence.
FEATURES
source Location/Qualifiers
1..132449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-328C17"
/clone_lib="RPCI-11.2"
1..1562
/note="L1M4 repeat: matches 4133. .5719 of consensus"
1563..1641
/note="L1M4 repeat: matches 1..80 of consensus"
1642..2199
/note="L1M4 repeat: matches 3396. .3937 of consensus"
```



```
repeat_region 2200..2474
/note="AluX repeat: matches 38..312 of consensus"
repeat_region 2475..4279
/note="L1M4 repeat: matches 1686..3396 of consensus"
repeat_region 4306..4589
/note="L1M4 repeat: matches 1254..1544 of consensus"
misc_feature complement(4704..5324)
/note="match: GSS: Em:AQ628183"
repeat_region 4799..4938
/note="L1M4C repeat: matches 785..924 of consensus"
misc_feature complement(4809..5326)
/note="match: GSS: Em:AQ774852"
misc_feature complement(4860..5272)
/note="match: GSS: Em:AQ787689"
repeat_region 4943..5094
/note="L1M4C repeat: matches 272..420 of consensus"
misc_feature complement(5020..5329)
/note="match: GSS: Em:AQ494601"
repeat_region 5142..5435
/note="L1M1D repeat: matches 170..428 of consensus"
repeat_region 5436..5744
/note="AluX repeat: matches 1..310 of consensus"
repeat_region 5745..5869
/note="L1M1D repeat: matches 428..568 of consensus"
repeat_region 6450..6686
/note="MIR repeat: matches 7..261 of consensus"
repeat_region 6916..7994
/note="L1M4 repeat: matches 3582..4688 of consensus"
repeat_region 7984..8050
/note="L1M8 repeat: matches 5808..5884 of consensus"
repeat_region 8064..8169
/note="L1M4 repeat: matches 3479..3593 of consensus"
repeat_region 8178..8224
/note="L1M84 repeat: matches 6124..6169 of consensus"
repeat_region 8225..8525
/note="AluX repeat: matches 1..301 of consensus"
repeat_region 8526..8915
/note="L1M84 repeat: matches 5725..6124 of consensus"
misc_feature complement(8862..9353)
repeat_region 9029..9150
/note="match: GSS: Em:AQ355497"
misc_feature complement(9052..9484)
/note="match: GSS: Em:AQ175845"
misc_feature 9388..9675
/note="match: GSS: Em:AQ135206"
misc_feature 9487..9840
/note="match: GSS: Em:AQ717684"
misc_feature 9491..9932
/note="match: GSS: Em:AQ147759"
repeat_region 9524..9697
/note="MER53 repeat: matches 1..189 of consensus"
repeat_region 10046..10143
/note="2 copies 49 mer 91% conserved"
repeat_region 11168..11361
/note="L1M1B repeat: matches 101..309 of consensus"
repeat_region 11448..11756
/note="AluY repeat: matches 1..309 of consensus"
repeat_region 12109..12515
/note="L2 repeat: matches 1747..2155 of consensus"
repeat_region 12798..12947
/note="MIR repeat: matches 29..206 of consensus"
repeat_region 12975..13026
/note="MIR repeat: matches 186..237 of consensus"
repeat_region 13005..13103
/note="L2 repeat: matches 2417..2516 of consensus"
repeat_region 13331..13664
/note="L1M1B repeat: matches 4..361 of consensus"
repeat_region 13704..13907
/note="L1R33 repeat: matches 297..511 of consensus"
repeat_region 15450..15530
/note="L2 repeat: matches 2668..2749 of consensus"
repeat_region 15876..15935
```

```
repeat_region 16202..16294
/note="MER81 repeat: matches 2..94 of consensus"
misc_feature 17335..17668
/note="match: GSS: Em:AQ104665"
repeat_region 17419..17715
/note="AluX repeat: matches 1..296 of consensus"
misc_feature 17995..18362
/note="match: GSS: Em:AQ151578"
repeat_region 19374..19496
/note="L2 repeat: matches 2579..2709 of consensus"
repeat_region 20271..20430
/note="MER20 repeat: matches 55..218 of consensus"
repeat_region 21095..21389
/note="AluX repeat: matches 1..295 of consensus"
repeat_region 21830..21946
/note="MIR repeat: matches 42..152 of consensus"
repeat_region 22328..22772
/note="L2 repeat: matches 2242..2710 of consensus"
repeat_region 22768..22848
/note="L2 repeat: matches 2661..2750 of consensus"
misc_feature 23314..24204
/note="CpG island"
/evidence=not_experimental
misc_feature 23596..23601
/note="Weak data."
repeat_region 25664..25783
/note="L2 repeat: matches 2588..2741 of consensus"
repeat_region 25786..26131
/note="MER33 repeat: matches 1..324 of consensus"
repeat_region 26858..27183
/note="AluY repeat: matches 1..301 of consensus"
repeat_region 27208..27334
/note="MIR repeat: matches 81..221 of consensus"
repeat_region 27673..27833
/note="L2 repeat: matches 7..175 of consensus"
misc_feature complement(28112..28516)
/note="match: GSS: Em:AQ212338"
misc_feature 28535..29088
/note="match: GSS: Em:B51862"
repeat_region 28560..28684
/note="MIR repeat: matches 79..210 of consensus"
misc_feature 28588..28828
/note="match: GSS: Em:B78941"
misc_feature 28588..28841
/note="match: GSS: Em:B88272"
repeat_region 28849..29048
/note="MER20 repeat: matches 2..214 of consensus"
repeat_region 30148..30232
/note="L2 repeat: matches 2610..2692 of consensus"
misc_feature complement(31265..31862)
/note="match: GSS: Em:AQ625926"
misc_feature complement(31333..31862)
/note="match: GSS: Em:AQ625934"
misc_feature complement(31491..31857)
/note="match: GSS: Em:AQ131221"
repeat_region 31733..31875
/note="L2 repeat: matches 2039..2197 of consensus"
repeat_region 33899..33992
/note="L2 repeat: matches 2578..2667 of consensus"
repeat_region 34111..34427
/note="MER31A repeat: matches 4..477 of consensus"
repeat_region 34427..35300
/note="MER5A repeat: matches 121..189 of consensus"
repeat_region 349..35435
```

Query Match 7.7%; Score 35.8; DB 90; Length 132449;  
Best Local Similarity: 52.3%; Pred.No. 7.4;  
Matches 79; Gaps: 0; Mismatches 72; Indels 0; Gaps 0;

QY 214 ggcctgga  
|||||  
|||ACACTTAGGAAGAGGATGTATGCCCTGGCTGCTGGACACT 115633  
DB 115574 GGCTCTG

Qy	274	gagatgccctagatcatcaccctaccacacccccctgacccaaccttggcccttaataaaa	333
Db	115634	GTGGGAACCTCTGGGATCCCTCCTCTAAAGGCTGTGCTACCTGATTTTGGAAACAATA	115693
Qy	334	cctatcttgcatttcgggaataaaccttcc	364
Db	115694	CAGAGGCTTCTCTCCTGCAAGAAACATGCC	115724
RESULT	11		
AC058816			
LOCUS	AC058816	171350 bp	DNA
DEFINITION	Homo sapiens chromosome 6 clone RP11-233K4, WORKING DRAFT SEQUENCE, 21 unordered pieces.	HTG	17-AUG-2000
VERSION	AC058816		
ACCESSION	AC058816.3	GI:9838113	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
R	1	(bases 1 to 171350)	
ENC	Waterston, R.H.		
HORS	The sequence of Homo sapiens clone		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 171350)		
REFERENCE	Waterston, R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (19-APR-2000) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On Aug 17, 2000 this sequence version replaced gi:7709945.		
	----- Genome Center -----		
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site: <a href="http://genome.wustl.edu/gsc/index.shtml">http://genome.wustl.edu/gsc/index.shtml</a>		
	----- Project Information -----		
	Center project name: H_NH0233K04		
	----- Summary Statistics -----		
	Sequencing vector: M13; 100%		
	Sequencing vector: plasmid; 0%		
	Chemistry: Dye-primer ET; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 0% of reads		
	Assembly program: Phrap; version 0.990319		
	Consensus quality: 161421 bases at least Q40		
	Consensus quality: 164672 bases at least Q30		
	Consensus quality: 166316 bases at least Q20		
	Insert size: 187000; agarose-fp		
	Insert size: 170998; sum-of-contigs		
	Quality coverage: 4.43 in Q20 bases; agarose-fp		
	Quality coverage: 4.92 in Q20 bases; sum-of-contigs		
	----- NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----		
	1	2122: contig of 2122 bp in length	
	2123	2222: gap of unknown length	
	2223	3812: contig of 1590 bp in length	
	3813	3912: gap of unknown length	
	3913	5636: contig of 1724 bp in length	
	5637	5736: gap of unknown length	
	5737	8302: contig of 2566 bp in length	
	8303	8402: gap of unknown length	
	8403	13037: contig of 4635 bp in length	
	13038	13137: gap of unknown length	
	13138	16443: contig of 3306 bp in length	

```

misc_feature      vector_side:left"
103476..119595
/note="assembly_name:Contig27"
119596..135009
/note="assembly_name:Contig28"
135110..152926
/note="assembly_name:Contig29"
153027..171350
/note="assembly_name:Contig30"
BASE COUNT      44302 a 39607 c 39553 g 45877 t 2011 others
ORIGIN

```

Query Match 7.7%; Score 35.8; DB 72; Length 171350;

Best Local Similarity 52.3%; Pred. No. 7.6; Indels 0; Gaps 0; Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```

C 214 ggcctgacgcgattctcttaccattgcgattaccgcgtatgcgccttcatacaaa 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164544 GGCTCTGCCGGTAGCATTAGGAAGAGGATGTATGCCCTGGCTGCTGGAGCACT 164603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 274 gaggatgccttagtataccctaccacaccccgtagcaaaccttggccttataataaaa 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164604 GTGGGAACCTGGGATCCCTCTCTAAGGCTGTGTACCTGATTGGAAACAATA 164663
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 cctattttgtcatttcggaataaaattttcc 364
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164664 CAGAGGCTTCTCTCCTGCAAGAAACATGCC 164694

```

# RESULT 12

```

AC044842 AC044842 172307 bp DNA HTG 25-JUN-2000
LOCUS Homo sapiens chromosome 2 clone RP11-178E20 map 2, WORKING DRAFT
DEFINITION SEQUENCE, 23 unordered pieces.
ACCESSION AC044842
VERSION 2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome 2, clone RP11-178E20
2 (bases 1 to 172307)
Anderson.S., Baldwin.J., Barna.N., Bastien.V., Bede.F.,
Boguslavsky.L., Boukhgalter.B., Brown.A., Burkett.G.,
Campopiano.A., Castle.A., Choepel.P., Colangelo.M., Collins.S.,
Collimore.A., Cooke.P., Dearellano.K., Dewar.K., Diaz.J.S.,
Dodge.S., Domino.M., Doyle.M., Ferreira.P., FitzHugh.W., Gage.D.,
Galagan.J., Gardyna.S., Ginde.S., Goyette.M., Graham.L.,
Grand-Pierre.N., Grant.G., Hagos.B., Heaford.A., Horton.L.,
Howland.J.C., Iliev.I., Johnson.R., Jones.C., Kann.L., Karatas.A.,
Klein.J., Lacombe.K., Lamazares.R., Landers.T., Lehoczyk.J.,
Levine.R., Lieu.C., Liu.G., Locke.K., Macdonald.P., Marquis.N.,
McCarthy.M., McEwan.P., McGurk.A., McKernan.K., McPheeters.R.,
Meldrum.J., Meneus.L., Mihova.T., Miranda.C., Mlenka.V., Morrow.J.,
Murphy.T., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,
O'Neill.D., Olivier.T.M., Oliver.J., Peterson.K., Pierre.N.,
Pisani.C., Pollara.V., Raymond.C., Riley.R., Rogov.P., Rothman.D.,
Roy.A., Santos.R., Schauer.S., Severi.P., Spencer.B.,
Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,
Tessaye.S., Theodore.J., Tirrell.A., Travers.M., Trigilio.J.,
Vassiliev.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,
Young.G., Zainoun.J., Zimmer.A. and Zody.M.
Direct Submission
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7543808.
All repeats were identified using RepeatMasker.

```

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L9002  
 Center clone name: 178\_E20  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 156889 bases at least Q40  
 Consensus quality: 164477 bases at least Q30  
 Consensus quality: 167863 bases at least Q20  
 Insert size: 171000; agarose-fp  
 Insert size: 170107; sum-of-contigs  
 Quality coverage: 3.7 in Q20 bases; agarose-fp  
 Quality coverage: 3.7 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 23 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 19: contig of 19 bp in length
20 119: gap of 100 bp
120 1921: contig of 1802 bp in length
1922 2021: gap of 100 bp
2022 3689: contig of 1668 bp in length
3690 3789: gap of 100 bp
3790 6531: contig of 2742 bp in length
6532 6631: gap of 100 bp
6632 8134: contig of 1503 bp in length
8135 8234: gap of 100 bp
8235 11919: contig of 3685 bp in length
11920 12019: gap of 100 bp
12020 15244: contig of 3225 bp in length
15245 15344: gap of 100 bp
15345 17344: contig of 2000 bp in length
17345 17444: gap of 100 bp
17445 21830: contig of 4386 bp in length
21831 21930: gap of 100 bp
21931 25081: contig of 3151 bp in length
25082 25181: gap of 100 bp
25182 29736: contig of 4535 bp in length
29737 29836: gap of 100 bp
29837 36853: contig of 7017 bp in length
36854 36953: gap of 100 bp
36954 41270: contig of 4317 bp in length
41271 41370: gap of 100 bp
41371 46765: contig of 5395 bp in length
46766 46865: gap of 100 bp
46866 57601: contig of 10736 bp in length
57602 57701: gap of 100 bp
57702 67879: contig of 10178 bp in length
67880 67979: gap of 100 bp
67980 78335: contig of 10256 bp in length
78336 78335: gap of 100 bp
78336 89176: contig of 10841 bp in length
89177 89276: gap of 100 bp
89277 106081: contig of 16805 bp in length
106082 1181: gap of 100 bp
1182 121487: contig of 15206 bp in length
121488 121487: gap of 100 bp
121488 137207: contig of 15720 bp in length
137208 137307: gap of 100 bp
137308 153745: contig of 16438 bp in length

```







Search completed: November 7, 2001, 04:14:09  
Job time: 2979 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 03:25:40 ; Search time 130.19 Seconds  
(without alignments)  
2242.678 Million cell updates/sec

Title: US-09-521-640-2  
Perfect score: 465  
Sequence: 1 gggagatcgtgagttcac.....cccttctgttcaaacacn 465

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	7.3	549	AAC95226	Cat flea head and
2	32.4	7.0	607	AAC68823	Human head/neck tu
3	31.8	6.8	89047	22 AAF28547	Genomic fragment #
4	31.6	6.8	963	21 AAA98219	Human retrovirus L
5	30.6	6.6	50000	21 AAA96367	Polymorphic repeat
6	30.2	6.5	1166	20 AAZ19440	M. tuberculosis an
7	30.2	6.5	1166	20 AAZ19228	M. tuberculosis re
8	30	6.5	684	21 AAC52130	Arabidopsis thalia
9	30	6.5	757	21 AAA02266	Human colon cancer
10	30	6.5	1233	21 AAC38681	Arabidopsis thalia
11	30	6.5	1329	21 AAC34844	Arabidopsis thalia

C	12	30	6.5	5616	22	AAF63962	Human tankyrase1 3
	13	29.6	6.4	798	21	AAF53576	Arabidopsis thalia
	14	29.6	6.4	1095	21	AAC77435	Human ORFX ORF2990
	15	29.4	6.3	640	21	AAC95117	Cat flea hindgut a
	16	29.4	6.3	743	21	AAA01800	Human colon cancer
C	17	29.4	6.3	1053	20	AAX86032	Nucleic acid encod
	18	29.4	6.3	1053	20	AAX86033	Membrane penetrati
	19	29.4	6.3	1475	19	AAV43795	Rodent chemokine r
	20	29.4	6.3	9289	20	AAX34655	Hexaploid wheat DB
C	21	29.4	6.3	910715	20	AAX20248	Borrelia burgdorfe
	22	28.8	6.2	477	18	AAX83232	Breast cancer tumo
	23	28.8	6.2	477	19	AAV68830	DNA molecule encod
	24	28.8	6.2	477	21	AAC80755	Human breast tumo
C	25	28.6	6.2	383	19	AAV61979	Human mu-opioid re
	26	28.6	6.2	1116	20	AAZ18998	Human gene express
	27	28.6	6.2	3336	14	AAQ64652	Human Mannose-Bind
	28	28.6	6.2	3592	14	AAQ53529	Human Mannose-Bind
	29	28.6	6.2	3605	20	AAZ07143	Human mannan-bind
	30	28.4	6.1	403	22	AAF66785	Novel human polynu
C	31	28.4	6.1	646	21	AAC76174	Human ORFX ORF1729
C	32	28.4	6.1	3444	16	AAZ05249	CryIF/cryIA(b) chl
	33	28.4	6.1	3444	16	AAZ05251	CryIF/cryIA(b) chl
	34	28.4	6.1	3444	17	AAT18701	CryIF/cryIA(b) chl
	35	28.4	6.1	3444	17	AAT18723	CryIF/cryIA(b) cod
	36	28.4	6.1	3444	19	AAV62080	Plasmid pMYC244 c
	37	28.4	6.1	3444	19	AAV62082	Plasmid pMYC2523 c
	38	28.4	6.1	3450	16	AAZ05269	CryIA(c)/cryIF/cry
	39	28.4	6.1	3450	17	AAT18721	CryIA(c)/cryIF/cry
	40	28.4	6.1	3450	19	AAV62079	Plasmid pMYC239 B
	41	28.4	6.1	3522	12	AAQ10182	Lepidopteran-activ
	42	28.4	6.1	3522	14	AAQ47291	Delta endotoxin ge
	43	28.4	6.1	3522	16	AAZ05270	CryIF toxin with 1
	44	28.4	6.1	3522	16	AAZ05250	CryIF/436 chimeric
C	45	28.4	6.1	3522	17	AAT18702	CryIF/436 chimeric

#### ALIGNMENTS

RESULT 1  
AAC95226  
ID AAC95226 standard; cDNA; 549 BP.  
AC AAC95226;  
XX  
XX  
XX 19-FEB-2001 (first entry)  
XX  
XX Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1721.  
XX  
XX Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;  
KW vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.  
XX Ctenocephalides felis.  
XX  
XX  
XX WO2000061621-A2.  
XX  
XX 19-OCT-2000.  
XX  
XX 07-APR-2000; 2000WO-US09437.  
XX  
XX 09-APR-1999; 99US-0128704.  
XX  
XX (HESK-) HESKA  
XX  
XX Brandt KS, Gail. ; PJ, Stinchcomb DT, Wisniewski N;  
XX  
XX WPI; 2000-6563.  
XX  
XX Flea Malpighian tubule and head and nerve cord tissue derived nucleic  
XX acids useful for the prevention, diagnosis and treatment of flea  
XX infestations  
XX  
XX Claim 26; Paq. 164pp; English.



XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic  
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue  
CC or head and nerve cord (HNC) tissue. The invention also relates to the  
CC encoded proteins. The invention additionally encompasses expression  
CC constructs, recombinant viruses and recombinant cells comprising the  
CC nucleic acids of the invention, recombinant production of the proteins,  
CC antibodies against the proteins, a method of identifying inhibitors of  
CC the proteins, and compositions comprising the inhibitors for  
CC administration to an animal. The nucleic acids, and the proteins they  
CC encode may be used in the prevention, treatment and diagnosis of diseases  
CC associated with flea infestations. For example, the nucleic acids may be  
CC used to produce an HMT or HNC protein according to standard recombinant  
CC DNA methodology by inserting the nucleic acids into a host cell and  
CC culturing the cell to express the protein. The HMT and HNC nucleic acids  
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect  
CC and quantitate the presence of cat flea or other homologous nucleic acid  
CC sequences in samples. They may also be used to study the expression and  
CC function of the proteins and their role in metabolism. The HMT and HNC  
CC proteins may be used as antigens in the production of specific  
CC antibodies, and in assays to identify modulators (agonists and  
CC antagonists) of HMT and/or HNC protein expression and activity. The  
CC anti-HMT/HNC protein antibodies and antagonists may also be used to  
CC downregulate protein expression and activity. The antibodies may also be  
CC used as diagnostic agents for detecting the presence of flea polypeptides  
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The  
CC present sequence represents a cat flea HNC cDNA of the invention.  
XX  
SQ Sequence 549 BP; 201 A; 102 C; 98 G; 148 T; 0 other;

Query Match 7.3%; Score 34; DB 21; Length 549;  
Best Local Similarity 88.1%; Pred. No. 0.13;  
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 93 gcgcgcttacacgcggactgagaaacctgcgtaccgcgttaa 134  
||||| ||||||| ||||| ||||||| |||||  
Db 508 gcgcgcttacacgcggactgagaaacctgcgtaccgcgttaa 549

RESULT 2  
AAC68823  
ID AAC68823 standard; cDNA; 607 BP.  
XX  
AC AAC68823;  
XX  
DT 21-FEB-2001 (first entry)  
XX  
DE Human head/neck tumour related protein partial coding sequence #20.  
KW Head tumour; neck tumour; lung cancer; vaccine; cancer therapy; ss.  
Q. Homo sapiens.  
XX  
PN WO200065053-A2.  
XX  
PD 02-NOV-2000.  
XX  
PF 21-APR-2000; 2000WO-US10687.  
XX  
PR 23-APR-1999; 99US-0130906.  
PR 20-APR-2000; 2000US-0533870.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Wang T, Dillon DC;  
XX  
DR WPI; 2000-687345/67.  
XX  
PT Novel polypeptides comprising immunogenic portion of head and neck  
PT tumour protein useful for treating, diagnosing and monitoring cancer  
PT such as head, neck and lung cancer -  
XX

PS Claim 3; Page 76; 77pp; English.  
XX  
CC The present invention relates to a number of nucleic acid sequences which  
CC encode proteins associated with head, neck and lung tumours. These  
CC tumours are often not diagnosed until they have spread, and, of those who  
CC survive, most must endure alterations in facial and neck appearance as  
CC well as changes in speech, sight, smell, chewing, swallowing and taste  
CC perception. The coding sequences given, and the proteins they encode, can  
CC be used in the diagnosis, treatment and vaccination against cancer,  
CC particularly papillary and follicular carcinomas, papillary tumours,  
CC follicular adenoma, parathyroid hyperplasia, parotid cancer, lip cancer,  
CC squamous cell cancer of the tongue, oral tongue cancers and larynx  
CC cancer.  
XX  
SQ Sequence 607 BP; 154 A; 159 C; 130 G; 141 T; 23 other;

Query Match 7.0%; Score 32.4; DB 21; Length 607;  
Best Local Similarity 74.6%; Pred. No. 0.48;  
Matches 53; Conservative 0; Mismatches 17; Indels 1; Gaps 1;  
QY 82 cttaagctactcgcgcgttacacgcggactgagaaacctgcgtaccgcgttaattcttc 141  
||| ||| ||||||| ||||||| ||||||| ||| ||||| |||  
Db 262 ctcaattactcgcgcgttacacgcggactgagaaacctgcgt-ccactaatcgcttn 320  
QY 142 acacatccccc 152  
||||||| |  
Db 321 acacatccctt 331

RESULT 3  
AAF28547/c  
ID AAF28547 standard; DNA; 89047 BP.  
XX  
AC AAF28547;  
XX  
DT 04-APR-2001 (first entry)  
XX  
DE Genomic fragment #34.  
XX  
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
KW bronchopulmonary; endocarditis; meningitis; ss.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WO200078968-A2.  
XX  
PD 28-DEC-2000.  
XX  
PF 16-JUN-2000; 2000WO-US16649.  
XX  
PR 18-JUN-1999; 99US-0140121.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Lagace RE, Patterson C, Berg KL;  
XX  
DR WPI; 2001-041427/05.  
XX  
PT Genomic library for identifying diagnostic and therapeutic  
PT compositions, and for identifying virulence factors, regulatory  
PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
PT acids -  
XX  
PS Claim 1; Page 324-345; 545pp; English.  
XX  
CC The present invention relates to a Moraxella catarrhalis genomic library  
CC comprising a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAF28554). The library has a number of uses described in the  
CC specification e.g. is useful for identifying diagnostic and therapeutic  
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
CC aerobic, gram-negative diplococcus, normally found among the bacterial  
CC flora of human upper airways. M. catarrhalis is known to cause acute,

CC localised infections such as otitis media, sinusitis and bronchopulmonary  
CC infection and life-threatening, systemic diseases including endocarditis  
CC and meningitis.

SQ Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other;

Query Match 6.8%; Score 31.8; DB 22; Length 89047;  
Best Local Similarity 50.3%; Pred. NO. 8.1;  
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 119 cctgcgtagccgctaattctttcacacatccctatcgcactggcggtatctcaatgacc 178  
 ||| | |||| | ||| |||| | ||| | ||| |  
 Db 49455 CTTAGCACCCAATGGGTCTTAAACACTGCGACGAGTACACGCCACGACGGATGTGC 49396

Qy . 179 accatcgccctccaaccatgccaaactaatggcaatggccctgatcgatttctcttacca 238

Db 395 CGCAGTGTGTGCAACACAGTCCAAGTTTTCAAGTGTACCGAACCGATACCACGCTTAG 49336

Qy 239 ttgtcggatttaccgcctatqccqcttcatcaaca 273

db 49335 GTGTAATTACACTCGCAAAATGCCGCATCATCA 49301

RESULT 4  
AAA98219  
ID AAA98219 standard: DNA: 963 BP.

AC AAA98219:

30-JAN-2001 (first entry)

Human retrovirus LTR DNA fragment 149-LTR = 120-LTR.

Cell-specific expression; tissue-specific expression; gene therapy; LTR; U3-R segment; long terminal repeat; retroviral expression vector; ds.

OS Human retrovirus.

PN WO200053789-A2.

14-SEP-2000.

09-MAR-2000: 2000WO-EP02064.

0-MAR-1999: 99DE-1010650

PA (GSF-) GSF FORSCHUNGSZENTRUM UMWELT &amp; GESUNDHEIT

PT Leib-Moesch C. Schoen U. Baust C.

WPT: 2000-587442/55

PT Retroviral expression vector, useful in gene therapy, contains a promoter from a human endogenous retrovirus to provide cell-specific expression -

PS Disclosure: Page 66: 67pp: German.

CC This invention describes a novel retroviral expression vector (A)

containing DNA sequences (2) for packaging vector RNA and for cell-specific expression of proteins or peptides encoding by heterologous DNA (11). The sequences controlling cell-specific expression contain a cell-specifically regulatable promoter region (P) from a human endogenous retrovirus (HERV) DNA sequence. The invention also describes (a) mRNA and

CC eukaryotic cells containing (A) in integrated form; (d) virions,  
CC containing a retroviral expression vector RNA derived from (A); (e) a  
CC method for producing the virions of (d); (f) a method for incorporating  
CC protein-encoding nucleic acid sequences into a eukaryotic cell by  
CC infection with the virions of (d); and (g) a retroviral vector system  
CC containing (A) and a packaging cell line, that contains at least one  
CC (recombinant) retrovirus construct that encodes for the packaging

proteins of (A). (A) are used for cell- or tissue-specific expression of foreign genes *in vivo*. (B) are used to produce viruses for introducing into the chromosomal DNA of eukaryotic cells, preferably mammalian and specifically human. (A) retain the advantages of usual retroviral promoters with all the signal structures required for transcription in a small region within the U3-R segment, but without their disadvantages (excessive strength and limited cell specificity). Since (A) are derived from endogenous (harmless) viral sequences, they do not introduce any new viral sequences into the genome and recombination will not create new types of retrovirus. The promoters provide cell or tissue specific expression, according to which HERV they are derived from.

Sequence 963 BP: 241 A: 223 C: 218 G: 281 T: 0 other:

```
Query Match      6.8%; Score 31.6; DB 21; Length 963;
Best Local Similarity 56.3%; Pred. No. 1.1;
Matches 58: Conservative 0; Mismatches 45; Indels 0; Gaps 0
```

Qy 303 ccccgtagccaaccttgcccttaataaaacctaattgtcatttcgggaataaacttt 362

Qy . 363 cctqtctttcaqcaaatqaaqaacccctqctcnqtccctttqqq 405

Db 908 cttgtctctgcacacagggagaaacacctgctaagccccgtagg 950

## RESULT 5

AAA90307  
ID AAA96367 standard: DNA: 50000 BP.

AAA  
AC  
AAA96367;

DT 08-FEB-2001 (first entry)

Polymorphic repeat microsatellite sequences present in the CTLA4 locus.

Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene; ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CCRL; lupus; KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy; KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma; KW thyroiditis; postpartum thyroiditis; rheumatoid arthritis; KW Hashimoto's disease; coeliac disease; ss.

OS Homo sapiens.

PN WO200056856-A2.

28-SEP-2000.

24-MAR-2000: 2000WO-US07938.

PR 25-MAR-1999: 99US-0126215.

PA (GEM) GENETICS INST INC.

PI Ling V, Wu P, Gray GS:

WPI: 2000-628257/60.

PT Determining predisposition of humans to develop autoimmune disease  
PT Involves detection of polymorphic microsatellite repeat sequence within  
PT human costimulatory receptor gene locus -

PS Disclosure: Page 138-142; 160pp; English;

Two human bacterial artificial chromosome (BAC) clones that included and flanked the human CTLA-4 locus were cloned and sequenced. The sequence data were assembled into a contiguous sequence that is presented in AA96363-68. AA96363-64 comprise BAC clone 22700, and AA96365-68 comprise BAC clone 22608. The sequences contain polymorphic microsatellite repeat (PMR) sequences. The specification describes a

CC. method for determining the predisposition of a human subject to develop  
 CC autoimmune disease. The method comprises detecting a PMR sequence in the  
 CC CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene  
 CC locus (hCGR). PMR sequences vary in length among individuals and can be  
 CC amplified to generate products that differ in size. These products can  
 CC then be detected by rapid and convenient high resolution processes. The  
 CC method is useful for determining the predisposition of insulin-dependent  
 CC diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune  
 CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,  
 CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,  
 CC coeliac disease and leprosy. PMR sequences within hCGR are useful as  
 CC markers in a variety of assays and in the field of forensic medicine,  
 CC disease diagnosis and human genome mapping.  
 XX  
 SQ Sequence 50000 BP; 14612 A; 9948 C; 10072 G; 15368 T; 0 other;

Query Match 6.6%; Score 30.6; DB 21; Length 50000;  
 Best Local Similarity 45.3%; Pred. No. 16; Indels 0; Gaps 0;  
 Matches 111; Conservative 0; Mismatches 134;

QY 14 agttactcaatattagcccttcagactcgtgttaagacctcgttcagccgaatccctt 73  
 D' 3472 attccttcagcttcagcctacccttccccaggaggttccaggaaacacacaaagcagg 26531  
 QY 74 tatgagcgttaagtactcgcgcgtttacacgcgactgagaaacctcgtgaccgccta 133  
 Db 26532 tctctgctcatgtatttattcttagttcttgtaacttattcctgcacatatctt 26591  
 QY 134 atctcttcacacatccctatccactgcactgcggtattctcaatgccaccatcgccttccaa 193  
 Db 26592 atacctcccaactctctcattagatctggtcagttaccatgcgccaccatgattgga 26651  
 QY 194 ccattgccaaactaatggcaatggccctgatcggattttcttaccattgctggatttacc 253  
 Db 26652 tgcatttgcattatgatattgtttggtgtgtgttcccccacccaaactctatgttgatt 26711  
 QY 254 gctat 258  
 Db 26712 gcaat 26716

RESULT 6  
 AAZ19440  
 ID AAZ19440 standard; cDNA; 1166 BP.  
 XX  
 AC AAZ19440;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis antigen 3' Erdsn-7 cDNA sequence.  
 X'  
 K' .Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test; ss.  
 XX

OS Mycobacterium tuberculosis.  
 XX  
 PN WO9942076-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03268.  
 XX  
 PR 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;  
 XX  
 DR WPI; 1999-527409/44.

XX  
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 XX skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Claim 4; Page 259; 299pp; English.  
 XX  
 CC The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAZ19083 to  
 CC AAZ19225 are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 1166 BP; 117 A; 371 C; 121 G; 207 T; 350 other;

Query Match 6.5%; Score 30.2; DB 20; Length 1166;  
 Best Local Similarity 28.2%; Pred. No. 3.6;  
 Matches 71; Conservative 46; Mismatches 135; Indels 0; Gaps 0;

QY 52 cctcggtcagcggaaatccctttatgagcgttaagtcaactgcgcggttacacgcggac 111  
 Db 797 cckmtctctctmckccymwcntcmkynccctcnmtckytccctcnmrycyvyak 856  
 QY 112 tgagaacactggtaccgcgttaattcttccacatccctcctcactgcgggtatctc 171  
 Db 857 cakcnctcccankmakctctctcccakmkacnckcccccctctcctcctcctc 916  
 QY 172 aatgaccacactgccttccacacatgccaaatggcaatggcctgatcggattctt 231  
 Db 917 wcyatctctctcwcnycmymkmacnckcyaytcnactmmwnccanctctctnct 976  
 QY 232 cttacattgctggatttaaccgctatggcgttctatcaacagaggatgccctagtatca 291  
 Db 977 cwckacgtyckckctmcknkmynmrwctyrcctckkccnccrcknmckmctmctcc 1036  
 QY 292 cctaccacacac 303  
 Db 1037 wmkmtcccwccc 1048

RESULT 7  
 AAZ19228  
 ID AAZ19228 standard; cDNA; 1166 BP.  
 XX  
 AC AAZ19228;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis recombinant antigen cDNA encoding 3' Erdsn-8.  
 XX  
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity; ss.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9942118-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03265.  
 XX  
 PR 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

```
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527416/44.
XX
XX New polypeptide comprising antigenic portions of M. tuberculosis
PT
XX Claim 4; Page 304; 323pp; English.
XX
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
XX Sequence 1166 BP; 117 A; 371 C; 121 G; 207 T; 350 other;
SQ
    Query Match          6.5%; Score 30.2; DB 20; Length 1166;
    Best Local Similarity 28.2%; Pred. No. 3.6;
    Matches 71; Conservative 46; Mismatches 135; Indels 0; Gaps 0;
QY 52 cctcggtccagcggaatccctttatgagcgcttaagtcaactgcgcggttacacgggac 111
Db 797 cckmtctctctmckcmwntccmkynccctccnmtcmktyctctcnmrycyayak 856
QY 112 tgagaacactgggtaccgccttaactctttcacacatccctatcgactggcggtatctc 171
Db 857 cakcmctcccckankcaactkctctcccckakmkcaacnckcyatcnactmmwncanctctctnct 916
QY 172 aatgaccacacatgccttccaaaccatgccaactaatggcaatggccctgatcgattct 231
Db 917 wctyatctckctcwnyymkmcaacnckcyatcnactmmwncanctctctnct 976
QY 232 cttacattgtggattaccgccttatggcgtttcatcaacagagatgcccttagtata 291
Db 977 cwcakctgtyckckcmckcnymcnmrwctyrotckkccnccrckmcmckmctctcc 1036
QY 292 cctaccaccacac 303
Db 1037 wmkmtccccc 1048
RESULT 8
AA* 130/C
IU AC52130 standard; DNA; 684 BP.
XX
XX AAC52130;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 70438.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130851.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
```

PR	19-JUL-1999;	9905-01443329;
PR	19-JUL-1999;	9905-01443333;
PR	19-JUL-1999;	9905-01443334;
PR	19-JUL-1999;	9905-01443335;
PR	20-JUL-1999;	9905-01443352;
PR	20-JUL-1999;	9905-01446322;
PR	20-JUL-1999;	9905-01448484;
PR	21-JUL-1999;	9905-01448484;
PR	21-JUL-1999;	9905-01450886;
PR	21-JUL-1999;	9905-01450888;
PR	22-JUL-1999;	9905-01450885;
PR	22-JUL-1999;	9905-01450887;
PR	22-JUL-1999;	9905-01450889;
PR	22-JUL-1999;	9905-01451922;
PR	23-JUL-1999;	9905-01451922;
PR	23-JUL-1999;	9905-01451922;
PR	23-JUL-1999;	9905-01452018;
PR	23-JUL-1999;	9905-01452244;
PR	26-JUL-1999;	9905-01452767;
PR	27-JUL-1999;	9905-01459118;
PR	27-JUL-1999;	9905-01459118;
PR	27-JUL-1999;	9905-01459519;
PR	28-JUL-1999;	9905-01459511;
Pi	12-AUG-1999;	9905-01463866;
Pk	02-AUG-1999;	9905-01463888;
PR	02-AUG-1999;	9905-01463889;
PR	03-AUG-1999;	9905-01470338;
PR	04-AUG-1999;	9905-01472004;
PR	04-AUG-1999;	9905-01473302;
PR	05-AUG-1999;	9905-01471922;
PR	06-AUG-1999;	9905-01472600;
PR	06-AUG-1999;	9905-01474116;
PR	09-AUG-1999;	9905-01474935;
PR	09-AUG-1999;	9905-01479335;
PR	10-AUG-1999;	9905-01481711;
PR	11-AUG-1999;	9905-01483319;
PR	12-AUG-1999;	9905-01483434;
PR	13-AUG-1999;	9905-01485655;
PR	13-AUG-1999;	9905-01486844;
PR	16-AUG-1999;	9905-01493688;
PR	17-AUG-1999;	9905-01491755;
PR	18-AUG-1999;	9905-01494226;
PR	20-AUG-1999;	9905-01497222;
PR	20-AUG-1999;	9905-01497222;
PR	20-AUG-1999;	9905-01499229;
PR	23-AUG-1999;	9905-01499022;
PR	23-AUG-1999;	9905-01499300;
PR	25-AUG-1999;	9905-01500566;
PR	26-AUG-1999;	9905-01500884;
PR	27-AUG-1999;	9905-01510665;
F	27-AUG-1999;	9905-01510665;
F	27-AUG-1999;	9905-01510665;
P.	27-AUG-1999;	9905-01510800;
PR	30-AUG-1999;	9905-01513303;
PR	31-AUG-1999;	9905-01514338;
PR	01-SEP-1999;	9905-01519300;
PR	07-SEP-1999;	9905-01523663;
PR	10-SEP-1999;	9905-01530700;
PR	13-SEP-1999;	9905-01537558;
PR	13-SEP-1999;	9905-01540118;
PR	16-SEP-1999;	9905-01540339;
PR	20-SEP-1999;	9905-01547779;
PR	22-SEP-1999;	9905-01551339;
PR	23-SEP-1999;	9905-01554866;
PR	24-SEP-1999;	9905-01556659;
PR	28-SEP-1999;	9905-01564558;
PR	29-SEP-1999;	9905-01565696;
PR	04-OCT-1999;	9905-01571177;
PR	05-OCT-1999;	9905-01577553;
PR	06-OCT-1999;	9905-01578665;
PR	07-OCT-1999;	9905-01580229;
PR	08-OCT-1999;	9905-01582322;
PR	12-OCT-1999;	9905-01583669;
PR	13-OCT-1999;	9905-01592923;

PR	16-APR-1999;	99US-0128645.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134766.
PR	19-MAY-1999;	99US-0134944.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139819.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	08-JUL-1999;	99US-0141280.
PR	09-JUL-1999;	99US-0141287.
PR	12-JUL-1999;	99US-0141842.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143521.
PR	15-JUL-1999;	99US-0143624.
PR	16-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.

PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
P. 02-AUG-1999; 99US-0147038.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
F 30-AUG-1999; 99US-0151303.  
F 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 6.5%; Score 30; DB 21; Length 1233;  
Best Local Similarity 53.4%; Pred. No. 4.3;  
Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
  
Qy 165 gtatctcaatgaccaccatcgcttccaaaccatgccaactaatggcaatggccctgatcg 224  
          |||||  |||||||  |||||  |||||  |||||  |||||  |||||  |||||  
Db 215 GTTTCGATCGACCACTCCGACTCGATCCCAACCATCTCTCTCTATACATGATCT 156  
  
Qy 225 gattctcttaccattgctcgatttaccgcgctatgcgcttccatcacagagatgcc 282  
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  
Db 155 AAAATCTCTCCGAATTCGGAATCTCCCGTCACCGGAGATCCACCCCGTGAAAGCC 98  
  
RESULT 11  
AAC34844/c  
ID AAC34844 standard; DNA; 1329 BP.  
XX  
AC AAC34844;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 8091.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0129845.  
PR 21-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.

PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130691.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	10-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139459.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140823.	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140991.	PR	10-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141842.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0142154.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142390.	PR	22-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142803.	PR	23-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142920.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0143542.	PR	29-SEP-1999;	99US-0156596.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157117.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157753.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159330.
PR			PR	14-OCT-1999;	99US-0159331.





PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 7-JUN-1999; 99US-0138540.  
PR 12-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 6-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.



[illegible]

Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;

WPI; 2000-656323/63.

Flea Malpighian tubule and head and nerve cord tissue derived nucleic acids useful for the prevention, diagnosis and treatment of flea infestations -

Claim 26; Page 784; 964pp; English.

The invention relates to novel cat flea (*Ctenocephalides felis*) nucleic acids which are expressed in hindgut and Malpighian tubule (HMT) tissue or head and nerve cord (HNC) tissue. The invention also relates to the encoded proteins. The invention additionally encompasses expression constructs, recombinant viruses and recombinant cells comprising the nucleic acids of the invention, recombinant production of the proteins, antibodies against the proteins, a method of identifying inhibitors of the proteins, and compositions comprising the inhibitors for administration to an animal. The nucleic acids, and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with flea infestations. For example, the nucleic acids may be used to produce an HMT or HNC protein according to standard recombinant DNA methodology by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The HMT and HNC nucleic acids may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect and quantitate the presence of cat flea or other homologous nucleic acid sequences in samples. They may also be used to study the expression and function of the proteins and their role in metabolism. The HMT and HNC proteins may be used as antigens in the production of specific antibodies, and in assays to identify modulators (agonists and antagonists) of HMT and/or HNC protein expression and activity. The anti-HMT/HNC protein antibodies and antagonists may also be used to downregulate protein expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of flea polypeptides in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The present sequence represents a cat flea HMT cDNA of the invention.

Sequence 640 BP: 208 A: 135 C: 133 G: 153 T: 11 other:

	Query Match	6.3%	Score 29.4	DB 21	Length 640
	Best Local Similarity	53.3%	Pred. No. 5.1		
	Matches 96	Conservative 0	Mismatches 80	Indels 4	Gaps 2
Q:	37	agactcgctaaagacactcggtccagccgaatccctttatgagcgcttaagtcaactgcgcg	96		
Db	452	aaactcgagggggccgggtcccaatcgccctatagtcgtagtcgtatacaattcaactggcg	511		
Qy	97	cgctttacagc--cggaactgagaaacctgctaccgcgctaatctctttcacacatccccctat	154		
Db	512	cgtttacagctcgctgctgggaaacctgcgtaccacacttatcgcttgnacacatccccctt	571		
Qy	155	cggaactggcggtatctctcaatgaccacacatcgctcttcccaacctgcgaactaatggcaatg	214		
Db	572	cnccatg--cgataagaaagcccnccgactcgcttttncacagctccanctgattgggaang	629		

Search completed: November 7, 2001, 04:11:03  
Job time: 2723 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 03:23:35 ; Search time 1170.27 Seconds  
(without alignments)  
3756.037 Million cell updates/sec

Title: US-09-521-640-2  
Perfect score: 465  
Sequence: 1 ggggagatagtgttcac.....cccttctgttcaaacacn 465

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Sequences: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est33:\*  
111: gb\_est34:\*  
112: gb\_est35:\*  
113: gb\_est36:\*  
114: gb\_est37:\*  
115: gb\_est38:\*  
116: gb\_est39:\*

117: gb\_est48: \*  
118: gb\_est49: \*  
119: gb\_est50: \*  
120: gb\_est51: \*  
121: gb\_est52: \*  
122: gb\_est53: \*  
123: gb\_est54: \*  
124: gb\_est55: \*  
125: gb\_est56: \*  
126: gb\_est57: \*  
127: gb\_est58: \*  
128: gb\_est59: \*  
129: gb\_est60: \*  
130: gb\_est61: \*  
131: gb\_est62: \*  
132: gb\_est63: \*  
133: gb\_est64: \*  
134: gb\_est65: \*  
135: gb\_est66: \*  
136: gb\_est67: \*  
137: gb\_est68: \*  
138: gb\_est69: \*  
139: gb\_est70: \*  
140: gb\_est71: \*  
141: gb\_est72: \*  
142: gb\_est73: \*  
143: gb\_est74: \*  
144: gb\_est75: \*  
145: gb\_est76: \*  
146: gb\_est77: \*  
147: gb\_est78: \*  
148: gb\_est79: \*  
149: gb\_est80: \*  
150: gb\_est81: \*  
151: gb\_est82: \*  
152: gb\_est83: \*  
153: gb\_est84: \*  
154: gb\_est85: \*  
155: gb\_est86: \*  
156: gb\_est87: \*  
157: gb\_est88: \*  
158: gb\_est89: \*  
159: gb\_est90: \*  
160: gb\_est91: \*  
161: gb\_est92: \*  
162: gb\_est93: \*  
163: gb\_est94: \*  
164: gb\_est95: \*  
165: gb\_est96: \*  
166: gb\_est97: \*  
167: gb\_est98: \*  
168: gb\_est99: \*  
169: gb\_est100: \*  
170: gb\_est101: \*  
171: gb\_est102: \*  
172: gb\_est103: \*  
173: gb\_est104: \*  
174: gb\_est105: \*  
175: gb\_est106: \*  
176: gb\_est107: \*  
177: gb\_est108: \*  
178: gb\_est109: \*  
179: gb\_est110: \*  
180: gb\_est111: \*  
181: gb\_est112: \*  
182: gb\_est113: \*  
183: gb\_est114: \*  
184: gb\_est115: \*  
185: gb\_est116: \*  
186: gb\_est117: \*  
187: gb\_est118: \*  
188: gb\_est119: \*  
189: gb\_est120: \*

190: gb\_est121: \*  
191: gb\_est122: \*  
192: gb\_est123: \*  
193: gb\_est124: \*  
194: gb\_est125: \*  
195: gb\_est126: \*  
196: gb\_est127: \*  
197: gb\_est128: \*  
198: gb\_est129: \*  
199: gb\_est130: \*  
200: gb\_est131: \*  
201: gb\_est132: \*  
202: gb\_est133: \*  
203: gb\_est134: \*  
204: gb\_est135: \*  
205: gb\_est136: \*  
206: gb\_est137: \*  
207: gb\_est138: \*  
208: gb\_est139: \*  
209: gb\_est140: \*  
210: gb\_est141: \*  
211: gb\_est142: \*  
212: gb\_est143: \*  
213: gb\_est144: \*  
214: gb\_est145: \*  
215: gb\_est146: \*  
216: gb\_est147: \*  
217: gb\_est148: \*  
218: gb\_est149: \*  
219: gb\_est150: \*  
220: gb\_est151: \*  
221: gb\_est152: \*  
222: gb\_est153: \*  
223: gb\_est154: \*  
224: gb\_est155: \*  
225: gb\_est156: \*  
226: gb\_est157: \*  
227: gb\_est158: \*  
228: gb\_est159: \*  
229: gb\_est160: \*  
230: gb\_est161: \*  
231: gb\_est162: \*  
232: gb\_est163: \*  
233: gb\_est164: \*  
234: gb\_est165: \*  
235: gb\_est166: \*  
236: gb\_est167: \*  
237: gb\_est168: \*  
238: gb\_est169: \*  
239: gb\_est170: \*  
240: gb\_est171: \*  
241: gb\_est172: \*  
242: gb\_est173: \*  
243: gb\_est174: \*  
244: gb\_est175: \*  
245: gb\_est176: \*  
246: gb\_est177: \*  
247: gb\_est178: \*  
248: gb\_est179: \*  
249: gb\_est180: \*  
250: gb\_est181: \*  
251: gb\_est182: \*  
252: gb\_est183: \*  
253: gb\_est184: \*  
254: gb\_est185: \*  
255: gb\_est186: \*  
256: gb\_est187: \*  
257: gb\_est188: \*  
258: gb\_est189: \*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	89.2	19.2	579	225	AQ210729	HS_2229_A
2	64	13.8	489	233	AQ761577	HS_3187_A
3	63.8	13.7	468	233	AQ806004	HS_3025_A
4	60.6	13.0	740	235	AQ914869	nbcb050K
5	53.2	11.5	397	225	AQ208121	HS_3220_A
6	50.2	10.8	1195	162	BE035400	M004D10 M
7	45.6	9.8	677	32	AV702190	AV702190
8	45.2	9.7	173	226	AQ304554	HS_3247_B
9	44.4	9.5	300	167	BE401707	CNW02BL02
10	43.6	9.4	205	233	AQ763115	HS_3161_A
11	43.2	9.3	736	32	AV721604	AV721604
12	43	9.2	946	32	AV726956	AV726956
13	42.8	9.2	168	234	AQ823152	HS_3186_B
14	42.6	9.2	659	32	AV705287	AV705287
15	42.4	9.1	312	256	B45121	HS_1060-B1
16	42.2	9.1	259	233	AQ775293	HS_3151_A
17	42.2	9.1	1101	219	CNS00LT2	AL078714 Drosophila
18	41.4	8.9	419	224	AQ116709	HS_3080_A
19	41.4	8.9	818	32	AV726386	AV726386
20	41	8.8	545	139	BE755460	209303 MA
21	40.6	8.7	982	15	A1068596	mgae0003b
22	40.4	8.7	403	233	AQ785487	HS_3063_A
23	40.4	8.7	714	104	A1965418	sc71e07.Y
24	40.4	8.7	847	32	AV726234	AV726234
25	40.2	8.6	981	15	A1068440	mgae0002b
26	40	8.6	146	235	AQ891463	HS_3111_B
27	40	8.6	432	224	AQ139216	HS_3090_A
28	40	8.6	961	152	EG343702	HVSMG000
29	39.6	8.5	824	112	AW155132	mg1e00020
30	39.6	8.5	870	32	AV726543	AV726543
31	39.4	8.5	920	32	AV727278	AV727278
32	39.2	8.4	819	153	EG418812	HVSMEX002
33	39	8.4	768	15	A1068832	mgae0004b
34	38.8	8.3	457	233	AQ775019	HS_3155_A
35	38.8	8.3	890	164	BE214026	HV_CEB000
36	38.8	8.3	985	162	BE034549	MK01B09 M
37	38.6	8.3	1200	110	AV761384	AV761384
38	38	8.2	929	152	EG343386	HVSMG000
39	38	8.2	972	106	AL576318	AL576318
40	37.8	8.1	287	233	AQ805901	HS_3235_A
41	37.8	8.1	1017	162	BE039690	OG02E04 O
42	37.6	8.1	430	226	AQ301842	HS_3195_A
43	37.6	8.1	945	164	BE195997	HVSMEX009
44	36.8	7.9	263	120	AW800405	MR2-UK006
45	36.8	7.9	802	32	AV725250	AV725250

ALIGNMENTS

RESULT 1  
AQ210729  
LOCUS HS\_2229\_AL\_C03\_MR\_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2229 Col=5 Row=E, DNA sequence.  
DEFINITION  
ACCESSION AQ210729.1 GI:3619698  
VERSION  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 579)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and

JOURNAL MEDLINE COMMENT

scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 2229 row: E column: 5  
Class: BAC ends  
High quality sequence stop: 579.  
Location/Qualifiers  
1..579  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"  
/sex="male"

FEATURES source

BASE COUNT 148 a 181 c 111 g 138 t 1 others  
ORIGIN  
Query Match 19.2%; Score 89.2; DB 225; Length 579;  
Best Local Similarity 68.5%; Pred. No. 4e-17;  
Matches 152; Conservative 0; Mismatches 68; Indels 2; Gaps 2;

LOCUS

AQ761577 489 bp DNA GSS  
HS\_3187\_AL\_G06\_MR\_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3187 Col=11 Row=M, DNA sequence.  
ACCESSION AQ761577.1 GI:5627380  
VERSION  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 489)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and

RESULT 2

AQ761577  
LOCUS HS\_3187\_AL\_G06\_MR\_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3187 Col=11 Row=M, DNA sequence.  
DEFINITION  
ACCESSION AQ761577.1 GI:5627380  
VERSION  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 489)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and





```
source
1. 740
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe0050K20r"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa. The
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

BASE COUNT 195 a 173 c 190 g 181 t 1 others  
ORIGIN

Query Match 13.0%; Score 60.6; DB 235; Length 740;  
Best Local Similarity 61.7%; Pred. No. 4e-08;  
Matches 113; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

QY 139 ttcacacatccctatcgcaactgagcggtatctcattgacacacatcgcttccaacatg 198  
Db 342 tgcagacatccctttccgagctggcggtatgacaaaggccgacgcttcacagtg 401

QY 199 ccaactaatggcaatggcctgacgtatctcttaccattgctgattaccgctat 258  
Dr 402 cgcactgatgggaaggcctgagcgaattttcttaccgattggcggattttacacgat 461

QY 259 ggcgtcttcatcacagagatgacccctagtagtaca-cctaccacacaccccgtagcaacct 317  
Db 462 ggctcttagacaaattgctgagcgatagtagcagcccgacacacgacgtagcacaacc 521

QY 318 tgg 320  
Db 522 ttg 524

RESULT 5  
A0208121  
LOCUS A0208121 397 bp DNA GSS 18-SEP-1998  
DEFINITION HS\_3220.A2.H07\_MR CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3220 Col=14 Row=O, DNA sequence.  
ACCESSION A0208121  
VERSION A0208121.1 GI:3620856  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 397)  
REFERENCE Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3220 row: O column: 14  
Class: BAC ends  
High quality sequence stop: 397.

FEATURES  
source  
1. 397  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3220 Col=14 Row=O"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"  
BASE COUNT 93 a 111 c 85 g 106 t 2 others  
ORIGIN

Query Match 11.5%; Score 53.6; DB 225; Length 397;  
Best Local Similarity 81.6%; Pred. No. 5.4e-06;  
Matches 62; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 178 caccatcgcttccaaccatgccaactaatggcaatggcctgacgattccttacc 237  
Db 162 ccgattcccttccaaaggcagctgattggcaatggcctgattccttacc 221

QY 238 attgtcggtattacc 253  
Db 222 attgtcggtattacc 237

RESULT 6  
BE035400  
LOCUS BE035400 1195 bp mRNA EST 07-JUN-2000  
DEFINITION M004D10 MO Mesembryanthemum crystallinum cDNA 5' similar to  
ribosomal protein 117, mRNA sequence.  
ACCESSION BE035400  
VERSION BE035400.1 GI:8330409  
KEYWORDS EST.  
SOURCE common ice plant.  
ORGANISM Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllidae; Caryophyllales; Alzooceae; Mesembryanthemum.  
1 (bases 1 to 1195)  
REFERENCE Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea  
H., Kavasaki,S., McColough,A., Michalowski,C.B., Palacio,C.,  
Scara,G., Wheeler,M. and Zepeda,G.R.  
TITLE Functional Genomics of Plant Stress Tolerance  
JOURNAL Unpublished (2000)  
COMMENT Contact: Michalowski,C.B.  
University of Arizona  
Bio Sciences West room 513, Tucson, AZ 85721, USA  
Tel: 520-621-7982  
Fax: 520-621-1697  
Email: michalowski@arizona.edu  
An open reading frame exists.  
Location/Qualifiers  
1..1195  
organism="Mesembryanthemum crystallinum"  
/db\_xref="taxon:3544"  
/clone\_lib="MO"  
/tissue\_type="apical meristem and leaf primordia"  
/dev\_stage="5 weeks"  
/note="no stress"

```

BASE COUNT      305 a      266 c      301 g      304 t      19 others
ORIGIN

Query Match      10.8%; Score 50.2; DB 162; Length 1195;
Best Local Similarity 66.4%; Pred. No. 8.3e-05;
Matches 87; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 88 tcaactccgcgttacacg-cggactgagaacactcgctaccgcgttaactcttcacaca 146
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 TAACTCCGCTGCTACACGCTCTGACTGGGAACCTCGTACCACACTAAATGCTTGAGACA 933
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 147 tccctatcgactcgcggtatctcaatgacacaccatcgctctcccaaccatgccaaactaa 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 TCCCTTTTGGCGTGGCGGAATTCGAAAGCCCGACGCGCTTTTCACAGTTGCAACTGAT 993
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 207 tggcaatggcc 217
|| | | | |
Db 994 TGCTATGGACC 1004

RESULT 7
AV702190      677 bp      mRNA      EST      08-OCT-2000
DE.-ITION AV702190 ADB Homo sapiens cDNA clone ADBCOH01 5', mRNA sequence.
ACCESSION AV702190
VERSION AV702190.1 GI:10718520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 677)
AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA ADB clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Source
1..677
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADBCOH01"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      166 a      161 c      156 g      194 t
ORIGIN

Query Match      9.8%; Score 45.6; DB 32; Length 677;
Best Local Similarity 66.7%; Pred. No. 0.002;
Matches 80; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 93 gcgcgcttacacg-cggactgagaacactcgctaccgcgttaactcttcacacatcccc 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 GCGCGCTTACAGCCTGACTGGAACCTTGCTGCTACCCACTAATCGCTTGAGACAATCCC 609
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 152 tatcgactcgcggtatctctcaatgacacaccatcgctctcccaaccatgccaaactaatggca 211
|| | | | |
Db 610 TTTCAGTTGCGAATAGCCAAAGGCGACCGATGGCTTTCCACAGTGGCAGCTGATGGCA 669
|| | | | |

```

```

RESULT 8
AQ304554      173 bp      DNA      GSS      16-DEC-1998
LOCUS HS_3247_B2_A10_MR_CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3247 Col-20 Row-B, DNA sequence.
ACCESSION AQ304554
VERSION AQ304554.1 GI:4024340
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 173)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3247 row: B column: 20
Class: BAC ends
High quality sequence stop: 173.
Location/Qualifiers
1..173
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3247 Col-20 Row-B"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/Note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT      53 a      53 c      32 g      35 t
ORIGIN

Query Match      9.7%; Score 45.2; DB 226; Length 173;
Best Local Similarity 66.7%; Pred. No. 0.0019;
Matches 80; Conservative 0; Mismatches 38; Indels 2; Gaps 1;

QY 152 tatcgactcgcggtatctcaa--tgaccaccatcgctctcccaaccatgccaaactaatgg 209
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53 TATCGCAGTGGCGATAAACCAAAATGCACGACCGATCCCTTCAACATTTCCGCACTGATG 112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 210 caatggccctgatcggtattctcttaccattgtcggatttaccgcgtatggcgcttcac 269
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 113 CAACGGCGCTGCGGATTTCTCTACCACGACGCGATCTACCCCATATAGAGCTTACTC 172
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
BE401707      300 bp      mRNA      EST      21-JUL-2000
LOCUS CNM02EL021 ITEC CNW Wheat Powdery Mildew Resistant Library Triticum
DEFINITION aestivum cDNA clone CNM02EL021, mRNA sequence.
ACCESSION BE401707
VERSION BE401707.1 GI:9361175
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 300)

```



/note=Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:					
xhoI"					
BASE COUNT	207 a	165 c	196 g	167 t	1 others
ORIGIN					
Query Match	9.3%; Score 43.2; DB 32; Length 736;				
Best Local Similarity	62.9%; Pred. No. 0.011;				
Matches	83; Conservative 0; Mismatches 48; Indels 1; Gaps 1;				
Qy	109	cagtgaagaactcgctaccgcctaattcttccacatccccctatcgcactgacgtgat	168		
Dd	541	GACTGGGAACCTGCGGTACCACATATTCGCTGAGACATCCCTTTCGCAGTTGCGGTATA	600		
Qy	169	ctcaatgaccacc-atcgctttccaacctgccaaactaatggcaatggccctgatcgat	227		
Dd	601	CGACAGGCCGCGCATGCCITCCACAGTGCCACCGAATCAAAATGAATGGAGCGTAT	660		
Qy	228	tctcttaccat	239		
Dd	661	ATTGTTCCCGT	672		
F T 12					
AVJ4956	AV726956	946 bp	mRNA	EST	17-OCT-2000
LOCUS					
DEFINITION	AV726956 HTC Homo sapiens cDNA clone HTCAVD01 5', mRNA sequence.				
ACCESSION	AV726956				
VERSION	AV726956.1 GI:10836377				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 946)				
AUTHORS	Gu,Y., Peng Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu, .S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu.G., Hu,R., Chen.J., Chen.Z. and Han.Z.				
TITLE	Homo sapiens cDNA HTC clones				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chnc.sh.cn This clone is available at CHGC in Shanghai.				
FEATURES	Location/Qualifiers				
source	1..946				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="HTCAVD01"				
	/clone_lib="HTC"				
	/tissue_type="Hypothalamus"				
	/dev_stage="Adult"				
	/lab_host="SOLR"				
	/note=Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:				
	xhoI"				
BASE COUNT	265 a	180 c	220 g	270 t	11 others
ORIGIN					
Query Match	9.2%; Score 43; DB 32; Length 946;				
Best Local Similarity	62.8%; Pred. No. 0.014;				
Matches	81; Conservative 0; Mismatches 47; Indels 1; Gaps 1;				
Qy	107	cgaactgagaaactcgctaccgcctaattcttccacatccccctatcgcactgacgtgat	166		
Dd	609	CGGACTGGAACCTCGGTACCAACTAATCGCTTGACATCCNCTTTGCAAGTGGGTAT	668		
Qy	167	atctcaatgaccacactcgcttcccacactgccaactgaatggcaatggcccttatcqa	226		

```

Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzachron@u.washington.edu
Sequence Tagged Connector
Plate: CT 782 row: N column: 1
Class: BAC ends
High quality sequence stop: 312.
Location/Qualifiers
    source
    1. .312
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="Plate=CT 782 Col=1 Row=N"
    /clone_lib="CIT Human Genomic Sperm Library C"
    /sex="M"
    /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
    E-Coli DH10B"
BASE COUNT      80 a      66 c      95 g      70 t      1 others
ORIGIN

Query Match      9.1%; Score 42.4; DB 256; Length 312;
Best Local Similarity 51.9%; Pred. No. 0.016;
Matches 94; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 161 gccgggtatctcaatgacacccatcgcccttccaaacatgccaaactaatggcaatggccctg 220
    || || || || || || || || || || || || || || || || || || || || ||
Db 226 GGGTTTACCCCGGTGACGGGATTGGCCTTACCCCAAGCCAGTTCTCGGAGGATCCTTC 167

Qy 221 atcggtattctcttaccattgtcgattaccgcgtatggcgttcatcaacagaggatg 280
    || || || || || || || || || || || || || || || || || || || || ||
Db 166 CTAGTTTTCCTACAGGATATTCCTTTTCCACATTAGACCTCATTTGTGCACAGAGATT 107

Qy 281 ccctagtatcaccctaccacacccccgtgaccaacctggcctttaaataaacctatct 340
    || || || || || || || || || || || || || || || || || || || || ||
Db 106 CCCTTCATAGATCCTCAAAATACAGTGTNTCCAAAGCTGCTCCATTAAAGAAAGCTTGCA 47

Qy 341 t 341
    |
Db 46 T 46

Search completed: November 7, 2001, 03:48:08
Job time: 1473 sec

```



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**